

## METHODS AND COMPOSITIONS FOR SIMULTANEOUS SACCHARIFICATION AND FERMENTATION

### Related Information

- 5        This application claims priority to U.S. provisional application number 60/214,137, entitled "Synergistic Hydrolysis of Carboxymethyl Cellulose and Acid Swollen Cellulose by Two Endoglucanases (EGZ and EGY)," filed June 26, 2000, and U.S. provisional application number 60/219,913, entitled "Methods and Compositions for Simultaneous Saccharification and Fermentation," filed July 21, 2000, both of which  
10      are incorporated herein in their entirety by this reference. The contents of all patents, patent applications, and references cited throughout this specification are hereby incorporated by reference in their entireties.

### Government Sponsored Research

- 15      This work was supported, in part, by grants from the U.S. Department of Agriculture, National Research Initiative (98-35504-6177 and 98-35505-6976), the U.S. Department of Energy, Office of Basic Energy Science (FG02-96ER20222), and the Florida Agricultural Experiment Station, University of Florida.

### Background of the Invention

- 20      Many environmental and societal benefits would result from the replacement of petroleum-based automotive fuels with renewable fuels obtained from plant materials (Lynd *et al.*, (1991) *Science* 251:1318-1323; Olson *et al.*, (1996) *Enzyme Microb. Technol.* 18:1-17; Wyman *et al.*, (1995) *Amer. Chem. Soc. Symp.* 618:272-290). Each  
25      year, the United States burns over 120 billion gallons of automotive fuel, roughly equivalent to the total amount of imported petroleum. The development of ethanol as a renewable alternative fuel has the potential to eliminate United States dependence on imported oil, improve the environment, and provide new employment (Sheehan, (1994) ACS Symposium Series No. 566, ACS Press, pp 1-53).

- 30      In theory, the solution to the problem of imported oil for automotive fuel appears quite simple. Rather than using petroleum, a finite resource, ethanol, a renewable resource, can be produced efficiently by the fermentation of plant material. Indeed, Brazil has demonstrated the feasibility of producing ethanol and the use of ethanol as a primary automotive fuel for more than 20 years. Similarly, the United States produces  
35      over 1.2 billion gallons of fuel ethanol each year. Currently, fuel ethanol is produced from corn starch or cane syrup utilizing either *Saccharomyces cerevisiae* or *Zymomonas mobilis* (*Z. mobilis*). However, neither of these sugar sources can supply the volumes needed to realize a replacement of petroleum-based automotive fuels. In addition, both

cane sugar and corn starch are relatively expensive starting materials, which have competing uses as food products.

Moreover, these sugar substrates represent only a fraction of the total carbohydrates in plants. Indeed, the majority of the carbohydrates in plants are in the form of lignocellulose, a complex structural polymer containing cellulose, hemicellulose, pectin, and lignin. Lignocellulose is found in, for example, the stems, leaves, hulls, husks, and cobs of plants. Hydrolysis of these polymers releases a mixture of neutral sugars including glucose, xylose, mannose, galactose, and arabinose. No known natural organism can rapidly and efficiently metabolize all of these sugars into ethanol.

Nonetheless, in an effort to exploit this substrate source, the Gulf Oil Company developed a method for the production of ethanol from cellulose using a yeast-based process termed simultaneous saccharification and fermentation (SSF) (Gauss *et al.* (1976) U.S.P.N. 3,990,944). Fungal cellulase preparations and yeasts were added to a slurry of the cellulosic substrate in a single vessel. Ethanol was produced concurrently during cellulose hydrolysis. However, Gulf's SSF process has some shortcomings. For example, fungal cellulases have been considered, thus far, to be too expensive for use in large scale bioethanol processes (Himmel *et al.*, (1997) Amer. Chem. Soc. pp. 2-45; Ingram *et al.*, (1987) *Appl. Environ. Microbiol.* 53:2420-2425; Okamoto *et al.*, (1994) *Appl. Microbiol. Biotechnol.* 42:563-568; Philippidis, G., (1994) Amer. Chem. Soc. pp. 188-217; Saito *et al.*, (1990) *J. Ferment. Bioeng.* 69:282-286; Sheehan, J., (1994) Amer. Chem. Soc. pp 1-52; Su *et al.*, (1993) *Biotechnol. Lett.* 15:979-984).

### Summary of the Invention

The development of inexpensive enzymatic methods for cellulose hydrolysis has great potential for improving the efficiency of substrate utilization and the economics of the saccharification and fermentation process. Accordingly, developing enzymes and, preferably, biocatalysts that produce such enzymes which can be used for the efficient depolymerization of a complex sugars and subsequent rapid fermentation of the sugar into alcohol, would be of great benefit.

Certain microbes, such as *Erwinia chrysanthemi*, produce a number of hydrolase and lyase enzymes, which are very effective in the degrading of plant tissues containing complex sugars. In particular, this organism produces two different endoglucanase activities (comprising EGY and EGZ) which have been discovered to function, when used in particular amounts, as highly effective enzyme compositions for degrading complex sugars. These enzymes may be used as crude extracts having a desired mixture of endoglucanase activity or, preferably, may be used as purified compositions.

Moreover, a biocatalyst, preferably a recombinant bacterium, more preferably a ethanologenic bacterium, can be engineered to express one or more of these enzymatic activities in particular amounts sufficient for degrading complex sugars. Such a biocatalyst is suitable for the efficient degradation of complex sugars and subsequent 5 fermentation into alcohol by a process known as simultaneous saccharification and fermentation (SSF). An advantage of the above endoglucanase compositions or biocatalysts is that the need for additional fungal cellulases for degrading the complex sugars is reduced or eliminated.

The present invention provides endoglucanase activities for carrying out the 10 degrading of a complex sugar and more preferably, the use of endoglucanase activities in particular ratios for optimal degrading of a complex sugar.

In addition, the invention provides recombinant host cells engineered for optimal expression and secretion of endoglucanase activities suitable for degrading complex sugars. Specifically exemplified are recombinant enteric bacteria, *Escherichia* and 15 *Klebsiella*, which express an endoglucanase under the transcriptional control of a surrogate promoter for optimal expression. In addition, also exemplified is a recombinant enteric bacterium that expresses two different endoglucanases *celY* and *celZ*, where each is under the transcriptional control of a surrogate promoter for optimal expression in a particular ratio.

20 The invention provides for the further modification of these hosts to include a secretory protein/s that allow for the increased production and/or secretion of the endoglucanases from the cell. In a preferred embodiment, the invention provides for the further modification of these hosts to include exogenous ethanologenic genes derived from an efficient ethanol producer, such as *Zymomonas mobilis*.

25 Accordingly, these hosts are capable of expressing high levels of proteins that may be used alone or in combination with other enzymes or recombinant hosts for the efficient production of alcohol from complex sugars.

More particularly, in a first aspect, the invention provides a composition for degrading an oligosaccharide containing, a first endoglucanase having a first degrading 30 activity, and a second endoglucanase having a second degrading activity, where the first and second degrading activities are present in a ratio such that the degrading of the oligosaccharide by the first and second endoglucanases is synergized.

In a second aspect, the invention provides a method for degrading an 35 oligosaccharide comprising, contacting an oligosaccharide with a first endoglucanase having a first degrading activity and a second endoglucanase having a second degrading activity, where the first and second degrading activities are present in a ratio such that

the degrading of the oligosaccharide by the first and second endoglucanases is synergized.

In one embodiment of the above aspects, the contacting of the oligosaccharide with the first endoglucanase and the second endoglucanase is performed in any order or 5 concurrently.

In one embodiment of the above aspects, the first endoglucanase or the second endoglucanase, or both the first and the second endoglucanases, are derived from a cell extract. The cell extract is derived from a bacterial cell, e.g., a bacterial cell that has been recombinantly engineered to express the first endoglucanase or the second 10 endoglucanase, or both the first and the second endoglucanases. In a related embodiment, the bacterial cell is selected from the family Enterobacteriaceae, and preferably, is either *Escherichia* or *Klebsiella*, and more preferably contains a first endoglucanase that is encoded by *celZ* and a second endoglucanase that is encoded by *celY*, and where *celZ* and *celY* are derived from *Erwinia*.

15 In another embodiment of the above aspects, the first endoglucanase is EGZ and the second endoglucanase is EGY, preferably in a ratio ranging from about 1:1 to, more preferably, about 9:1 to about 19:1.

In still another embodiment of the above aspects, the first endoglucanase or the second endoglucanase, or both the first and the second endoglucanase, are purified.

20 In even another embodiment of the above aspects, the degrading of an oligosaccharide is synergized by a factor ranging from about 1.1 to about 2.0, and preferably by about 1.8.

In yet another embodiment of the above aspects, the composition contains an additional enzyme, e.g., an endoglucanase, exoglucanase, cellobiohydrolase,  $\beta$ -25 glucosidase, endo-1,4- $\beta$ -xylanase,  $\alpha$ -xylosidase,  $\alpha$ -glucuronidase,  $\alpha$ -L-arabinofuranosidase, acetylersterase, acetylxylanesterase,  $\alpha$ -amylase,  $\beta$ -amylase, glucoamylase, pullulanase,  $\beta$ -glucanase, hemicellulase, arabinosidase, mannanase, pectin hydrolase, pectate lyase, or a combination thereof.

30 In a related embodiment of the above aspects, the additional enzyme is a glucanase derived from a fungus, preferably *T. longibranchiatum*.

In another related embodiment of the above aspects, the additional enzyme is an ethanologenic enzyme, preferably an ethanologenic enzyme such as pyruvate decarboxylase or alcohol dehydrogenase.

35 In another embodiment of the above aspects, the first endoglucanase and the second endoglucanase are packaged separately.

In another embodiment of the above aspects, the composition is used for simultaneous saccharification and fermentation.

In still another embodiment of the above aspects, the oligosaccharide is a celooligosaccharide, lignocellulose, hemicellulose, cellulose, pectin, or any combination thereof.

5 In still another embodiment of the above aspects, the composition or method is, respectively, used or conducted in an aqueous solution.

In a third aspect, the invention provides a recombinant host cell suitable for degrading an oligosaccharide containing a first heterologous polynucleotide segment encoding a first endoglucanase having a first degrading activity, where the segment is under the transcriptional control of a surrogate promoter; and a second heterologous 10 polynucleotide segment encoding a second endoglucanase having a second degrading activity, where the segment is under the transcriptional control of a surrogate promoter, and where the first endoglucanase and the second endoglucanase are expressed so that the first and the second degrading activities are present in a ratio such that the degrading of the oligosaccharide by the first and second endoglucanases is synergized.

15 In one embodiment, the first endoglucanase or the second endoglucanase, or both the first and the second endoglucanases are secreted. In a related embodiment, the first endoglucanase or the second endoglucanase, or both the first and the second endoglucanases, are derived from a cell extract. The cell extract is derived from a bacterial cell, e.g., a bacterial cell that has been recombinantly engineered to express the 20 first endoglucanase or the second endoglucanase, or both the first and the second endoglucanases. In a related embodiment, the bacterial cell is selected from the family Enterobacteriaceae, and preferably, is either *Escherichia* or *Klebsiella*, and more preferably *E. coli* B, *E. coli* DH5 $\alpha$ , or *Klebsiella oxytoca*.

25 In yet another embodiment of the above aspects, the composition contains an additional enzyme, e.g., an endoglucanase, exoglucanase, cellobiohydrolase,  $\beta$ -glucosidase, endo-1,4- $\beta$ -xylanase,  $\alpha$ -xylosidase,  $\alpha$ -glucuronidase,  $\alpha$ -L-arabinofuranosidase, acetylersterase, acetylxylanesterase,  $\alpha$ -amylase,  $\beta$ -amylase, glucoamylase, pullulanase,  $\beta$ -glucanase, hemicellulase, arabinosidase, mannanase, pectin hydrolase, pectate lyase, or a combination thereof.

30 In one embodiment, the first endoglucanase is encoded by *celZ* and the second endoglucanase is encoded by *celY*, and where *celZ* and *celY* are derived from *Erwinia*.

In a related embodiment of the above aspects, the additional enzyme is a glucanase derived from a fungus, preferably *T. longibranchiatum*.

35 In another related embodiment, the first endoglucanase is EGZ and the second endoglucanase is EGY.

In another related embodiment, the additional enzyme is a secretory enzyme, preferably a *pul* or *out* gene product.

DRAFT - NOT FOR FILING

*Draft B'* In another related embodiment of the above aspects, the host cell is ethanologenic, e.g., *E. coli* KO4 (ATCC 55123); *E. coli* KO11 (ATCC 55124), *E. coli* KO12 (ATCC 55125) and *E. coli* LY01 (ATCC \_\_\_\_\_), *K. oxytoca* M5A1, and *K. oxytoca* P2 (ATCC 55307).

- 5        In a fourth aspect, the invention provides a method for enhancing the degradation of an oligosaccharide by contacting an oligosaccharide with a host cell containing, a first heterologous polynucleotide segment encoding a first endoglucanase having a first degrading activity, where the segment is under the transcriptional control of a surrogate promoter; and a second heterologous polynucleotide segment containing a sequence  
10      encoding a second endoglucanase having a second degrading activity, where the segment is under the transcriptional control of a surrogate promoter. The method further provides that the first endoglucanase and the second endoglucanase are expressed so that the first and the second degrading activities are present in a ratio such that the degrading of the oligosaccharide by the first and second endoglucanases is synergized and thereby  
15      enhanced.

In one embodiment of the above aspect, the first endoglucanase or the second endoglucanase or both the first and the second endoglucanases are secreted.

In another embodiment, the host cell of the above method is ethanologenic.

In another embodiment, the method is conducted in an aqueous solution.

- 20      In even another embodiment, the method is used for simultaneous saccharification and fermentation.

In yet another embodiment, the method includes degrading an oligosaccharide selected from the group consisting of celooligosaccharide, lignocellulose, hemicellulose, cellulose, pectin, or any combination thereof.

- 25      In a fifth aspect, the invention provides a method of making a recombinant host cell suitable for degrading an oligosaccharide by introducing into the host cell a first heterologous polynucleotide segment encoding a first endoglucanase having a first degrading activity, where the segment is under the transcriptional control of a surrogate promoter; and a second heterologous polynucleotide segment containing a sequence  
30      encoding a second endoglucanase having a second degrading activity, where the segment is under the transcriptional control of a surrogate promoter. The method further provides that the first and second endoglucanases are expressed such that the first and the second degrading activities are present in a ratio such that the degrading of the oligosaccharide by the first and second endoglucanases is synergized.

- 35      In one embodiment of the above aspect, the first endoglucanase or the second endoglucanase or both the first and second endoglucanases are secreted.

In another embodiment, the host cell is ethanologenic.

In even another embodiment, the first endoglucanase is encoded by *celZ* and the second endoglucanase is encoded by *celY*, and *celZ* and *celY* are derived from *Erwinia*.

In still another embodiment, the surrogate promoter of the first heterologous polynucleotide segment or the second heterologous polynucleotide segment or both the 5 first and second polynucleotide segments, contains a polynucleotide fragment derived from *Zymomonas mobilis*.

In yet another embodiment, the recombinant host cell is suitable for simultaneous saccharification and fermentation, and preferably, is ethanologenic.

In a sixth aspect, the invention provides a method for making a recombinant host 10 cell integrant by introducing into the host cell a vector containing the polynucleotide sequence of pLOI2352 (SEQ ID NO: 17) and identifying a host cell having the vector stably integrated.

In a seventh aspect, the invention provides a method for expressing a 15 endoglucanase in a host cell by introducing into the host cell a vector containing the polynucleotide sequence of pLOI2306 (SEQ ID NO: 12) and identifying a host cell expressing the endoglucanase.

In an eighth aspect, the invention provides a method for producing ethanol from an oligosaccharide source by contacting the oligosaccharide source with a ethanologenic host cell containing a first heterologous polynucleotide segment encoding a first 20 endoglucanase having a first degrading activity, where the segment is under the transcriptional control of a surrogate promoter; and a second heterologous polynucleotide segment encoding a second endoglucanase having a second degrading activity, where the segment is under the transcriptional control of a surrogate promoter. The method further provides that the first and second endoglucanases are expressed so 25 that the first and the second degrading activities are present in a ratio such that the degrading of the oligosaccharide by the first and second endoglucanases is synergized resulting in a degraded oligosaccharide that is fermented into ethanol.

In one embodiment, the first endoglucanase is encoded by *celZ* and the second endoglucanase is encoded by *celY* gene, and *celZ* and *celY* are derived from *Erwinia*.

30 In another embodiment, the host cell further contains a heterologous polynucleotide segment encoding at least one *pul* gene or *out* gene.

*Sure B7* In even another embodiment, the host cell is selected from the family Enterobacteriaceae, preferably *Escherichia* or *Klebsiella*, more preferably *E. coli* KO4 (ATCC 55123), *E. coli* KO11 (ATCC 55124), *E. coli* KO12 (ATCC 55125), LY01 35 (ATCC \_\_\_\_\_), *K. oxytoca* M5A1, or *K. oxytoca* P2 (ATCC 55307).

In another embodiment, the method is conducted in an aqueous solution.

In still another embodiment, the oligosaccharide is selected from the group consisting of celooligosaccharide, lignocellulose, hemicellulose, cellulose, pectin, and any combination thereof.

5 In another embodiment, the heterologous polynucleotide segment is, or derived from, pLOI 2352 (SEQ ID NO: 17).

In yet another embodiment, the first endoglucanase is EGZ and the second endoglucanase is EGY.

10 In another embodiment, the surrogate promoter of the first polynucleotide segment or the second polynucleotide segment, or both the first and the second polynucleotide segments contains a polynucleotide fragment derived from *Zymomonas mobilis*.

15 In ninth aspect, the invention provides a vector containing the polynucleotide sequence of a plasmid, or fragment thereof, of pLOI2311, pLOI1620, pLOI2316, pLOI2317, pLOI2318, pLOI2319, pLOI2320, pLOI2323, pLOI2342, pLOI2348, pLOI2349, pLOI2350, pLOI2352, pLOI2353, pLOI2354, pLOI2355, pLOI2356, pLOI2357, pLOI2358, or pLOI2359.

20 In a tenth aspect, the invention provides a host cell containing a vector having the polynucleotide sequence of a plasmid, of fragment thereof, of pLOI2311, pLOI1620, pLOI2316, pLOI2317, pLOI2318, pLOI2319, pLOI2320, pLOI2323, pLOI2342, pLOI2348, pLOI2349, pLOI2350, pLOI2352, pLOI2353, pLOI2354, pLOI2355, pLOI2356, pLOI2357, pLOI2358, or pLOI2359.

In one embodiment, the host is *Klebsiella oxytoca* strain P2 (pCPP2006), *Klebsiella oxytoca* strain SZ6 (pCPP2006), *Klebsiella oxytoca* strain SZ21 (pCPP2006), or *Klebsiella oxytoca* strain SZ22 (pCPP2006).

25 In an eleventh aspect, the invention provides a method for degrading an oligosaccharide by obtaining a first endoglucanase having a first degrading activity, obtaining a second endoglucanase having a second degrading activity, and contacting an oligosaccharide with the first and second endoglucanases, where the first and second degrading activities are present in a ratio such that the degrading of the oligosaccharide by the first and second endoglucanases is synergized.

30 In a twelfth aspect, the invention provides a method for enhancing the degrading of an oligosaccharide by contacting an oligosaccharide with a first endoglucanase having a first degrading activity and a second endoglucanase having a second degrading activity, where the first and second degrading activities are present in a ratio such that 35 the degrading of the oligosaccharide by the first and second endoglucanases is synergized and thereby enhanced.

- In a related aspect, the invention provides a method for degrading, and/or for enhancing the degrading of, an oligosaccharide by contacting an oligosaccharide with a first endoglucanase having a first degrading activity and a second endoglucanase having a second degrading activity, where the first and second degrading activities are present in a ratio such that the degrading of the oligosaccharide by 5 the first and second endoglucanases results in a change in viscosity, preferably, a reduction in viscosity, more preferably by an amount of at least , e.g., 5 centipoise, 10 centipoise, 20 centipoise, 50 centipoise, 100 centipoise, 500 centipoise, or 1000 centipoise or more, or within a range thereof.
- In a preferred embodiment, the oligosaccharide is cellulose, e.g., amorphous cellulose or crystalline cellulose and may be from a source such as, e.g., paper, pulp, or plant fiber.
- 10 In a thirteenth aspect, the invention provides a recombinant host cell suitable for degrading an oligosaccharide containing a first heterologous polynucleotide segment encoding a first endoglucanase; and a second heterologous polynucleotide segment encoding a second endoglucanase.
- In a related aspect, the host cell is suitable for reducing the viscosity of an 15 oligosaccharide by comprising a first heterologous polynucleotide segment encoding a first endoglucanase; and a second heterologous polynucleotide segment encoding a second endoglucanase.
- In one embodiment, the first heterologous polynucleotide segment is under the transcriptional control of a surrogate promoter, and the second heterologous polynucleotide 20 segment is under the transcriptional control of a surrogate promoter.
- In another embodiment, the cell is a bacterial cell, preferably selected from the family Enterobacteriaceae, more preferably from the genus *Escherichia* or *Klebsiella*.
- In another embodiment, the first endoglucanase is encoded by *celZ* and the second endoglucanase is encoded by *celY*, and *celZ* and *celY* are derived from *Erwinia*.
- 25 In another embodiment, the first endoglucanase is EGZ and said second endoglucanase is EGY.
- Sub B3* In a fourteenth aspect, the invention provides a recombinant host strain of *Klebsiella oxytoca* strain P2 (pCPP2006) represented by a deposit with the American Type Culture Collection designated as deposit number ATCC \_\_\_\_\_.  
~~In a fifteenth aspect, the invention provides a recombinant host strain of *Klebsiella oxytoca* strain SZ6 (pCPP2006) represented by a deposit with the American Type Culture Collection designated as deposit number ATCC \_\_\_\_\_.  
In a sixteenth aspect, the invention provides a recombinant host strain of *Klebsiella oxytoca* strain SZ21 (pCPP2006) represented by a deposit with the American Type Culture Collection designated as deposit number ATCC \_\_\_\_\_.~~
- 30  
35

In a ~~seventeenth~~ aspect, the invention provides a recombinant host strain of *Klebsiella oxytoca* strain SZ22 (pCPP2006) represented by a deposit with the American Type Culture Collection designated as deposit number ATCC\_\_\_\_\_.

- In a ~~eighteenth~~ aspect, the invention provides a recombinant cell containing a
- 5 first heterologous polynucleotide segment encoding a first endoglucanase; and a second heterologous polynucleotide segment encoding a second endoglucanase, where the first polynucleotide segment encoding a first endoglucanase or the second polynucleotide segment encoding a second endoglucanase, or both are sufficiently homologous in an amino acid alignment to either the gene product of *celY* or *celZ* from *Erwinia* as to share
- 10 the functional activity of being capable of degrading a polysaccharide.

In ~~nineteenth~~ aspect, the invention provides an extract or composition derived from a host cell of the invention, e.g., a secreted polypeptide, a lysate or broth, or a pure, semi-pure, or unpurified enzymatic extract or polypeptide which is suitable for degrading and/or reducing the viscosity of a oligosaccharide when contacted thereto.

- 15 Other features and advantages of the invention will be apparent from the following detailed description and claims.

#### **Brief Description of the Drawings**

- Figure 1** shows fermentation rates for the ethanologenic recombinant host *E. coli* KO11 using rice hull substrates pretreated with dilute acid and supplemented with two different media.

- Figure 2** shows simultaneous saccharification and fermentation (SSF) rates for the ethanologenic recombinant host strain *K. oxytoca* P2 using mixed waste office paper. Insoluble residues from SSF were recycled as a source of bound cellulase enzymes and substrate during subsequent fermentations.

- Figure 3** shows the structure of the plasmid pLOI2171, a low copy promoter probe vector showing the orientation of the kanamycin resistance gene (*kan*) for selection, the temperature sensitive pSC101 replicon (Rep(ts)) for episomal maintenance of the plasmid, and the promoterless polysaccharase gene *celZ* encoding phospho-beta-glucosidase (EGZ).

- Figure 4** is a graph showing the high correspondence between the size of the zone of clearance on CMC indicator plates (x-axis) measured for a transformed bacterial colony and the amount of glucanase activity expressed (y-axis).

- Figure 5** shows the partial nucleotide sequence (SEQ ID NO: 1) of the *Z. mobilis* DNA fragment in the pLOI2183 plasmid that functions as a surrogate promoter. The full sequence has been assigned GenBank Accession Number AF109242 (SEQ ID NO: 2). Indicated are two transcriptional start sites (#), -35 and -10 regions, the Shine-

Delgarno site (**bold**), partial vector and *celZ* sequence (lowercase), and the *celZ* start codon (atg indicated in **bold**).

**Figure 6** represents electron micrographs of *E. coli* DH5 $\alpha$  cells harboring different plasmids expressing little if any (pUC19; panel A), moderate (pLOI2164; panel 5 B), and high levels (pLOI2307; panel C) of glucanase in the form of periplasmic inclusion bodies (pib) localized between the outer cell wall and the inner membrane (im). The bar shown represents 0.1  $\mu$ m.

**Figure 7** shows a schematic detailing the cloning strategy used to construct the 10 *celZ* integration vector pLOI2306, a genetic construct capable of being introduced into the genome of a recombinant host and conferring stable glucanase expression activity to the host.

**Figure 8** shows a schematic representation of the *celZ* integration vector 15 pLOI2306 (SEQ ID NO:12) with the locations of the surrogate promoter from *Z. mobilis*, the *celZ* gene from *E. chrysanthemi*, resistance markers (*bla* and *tet*), and *K. oxytoca* target sequence indicated.

**Figure 9** shows graphical depictions of the synergistic action of EGY and EGZ. Both enzymes were diluted to equal CMCase activities (1.5 IU/ml) with calculated synergies shown in parenthesis. Panel A shows a graph depicting the effect of enzyme ratios on synergy. Different amounts of EGY and EGZ were combined to maintain a 20 constant predicted activity (0.15 IU/ml) based on the contribution of individual enzymes. Assays were incubated with CMC for 1 hour at 35°C and terminated by boiling. Numbers on the X axis indicate the proportions of EGZ and EGY. Synergy is shown above each bar. Panel B shows a graph depicting the hydrolysis of CMC by EGZ and EGY, alone and in combination (9 parts EGZ + 1 part EGY). All assays contained 25 equal total activities (0.15 IU/ml) based on the sum of individual EGY and EGZ activities. Synergy is shown above each point for the combination of both enzymes. Panel C shows a graph depicting the hydrolysis of acid-swollen cellulose by EGZ and EGY, alone and in combination. A 9 to 1 ratio of EGZ to EGY was used for the 30 combined enzyme reaction. All assays contained 1.5 IU/ml based on the sum of individual EGY and EGZ activities. Synergy is shown above each point for the combination of both enzymes.

**Figure 10** represents a thin layer chromatography (TLC) analysis of the hydrolysis products from two complex sugars: acid-swollen cellulose and Avicel®. Approximately 1.5 IU and 25 IU of CMCase were used in reactions with acid-swollen 35 cellulose and Avicel®, respectively. Abbreviations for Y axis: G1, glucose; G2, cellobiose; G3, cellotriose; G4, cellotetraose; and G5, cellopentaose. Lanes: S, mixed celooligosaccharide standard; C, control lacking enzyme; Z, EGZ; Y, EGY, and Z + Y:

EGZ + EGY. Panel A shows results with acid-swollen cellulose after a 6-hour incubation with CMCCase (1  $\mu$ l loading). Panel B shows results with acid-swollen cellulose after a 6-hour incubation with CMCCase (2  $\mu$ l loading). Panel C shows results with Avicel<sup>®</sup> after a 48-hour incubation with CMCCase (10  $\mu$ l loading).

- 5      **Figure 11** represents a TLC analysis showing the hydrolysis of celooligosaccharides by EGZ and EGY. Each test contained approximately 0.07 IU of CMCCase per ml (2 hour incubation, 35°C). Abbreviations: S, mixed celooligosaccharides standard; G1, glucose; G2, cellobiose; G3, cellotriose; G4, cellotetraose; and G5, cellopentaose. Panel A shows the substrates before hydrolysis,
- 10     Panel B shows the substrates after incubation with EGY, Panel C shows the substrates after incubation with EGZ, and Panel D shows EGZ hydrolysis of cellopentaose after different periods of incubation (0, 5, 10, and 25).

- 15     **Figure 12** is a model illustrating the utilization of amorphous cellulose by *E. chrysanthemi*. Three glucosidases are used for the catabolism of amorphous cellulose. Two of these, EGY and EGZ are extracellular endoglucanases, which function, together in a synergistic fashion. EGY requires large substrate molecules and hydrolyzes these into shorter, insoluble fragments. EGY does not hydrolyze soluble celooligosaccharides (2 to 5 glucosyl residues). EGZ readily hydrolyzes soluble celooligosaccharides (cellopentaose and cellotetraose) and amorphous fragments of intermediate length to produce cellobiose and cellotriose. Cellobiose and cellotriose are phosphorylated during cellular uptake by a phosphoenolpyruvate-dependent phosphotransferase system. Hydrolysis is completed intracellularly by a third enzyme, phospho- $\beta$ -glucosidase. Resulting monomeric products (glucose and glucose-6-phosphate) are metabolized by glycolysis.

- 25     **Figure 13** is a schematic representation of the construction of a promoter-probe vector for *celY*. *Sau3AI* fragments of *Z. mobilis* chromosomal DNA were ligated into the *BamHI* site of pLOI2317 to provide a strong, surrogate promoter for *celY* coding region (solid segments). *Z. mobilis* DNA fragments (promoter 1 and promoter 2) are shown as open segments. Replicons and antibiotic resistance genes are stippled; other vector DNA is shown as thin connecting lines. Arrows indicate direction of transcription.

**Figure 14** is a depiction of transcriptional initiation sites and putative promoter regions for the *celY* promoter in DH5 $\alpha$  (pLOI2323).

- 35     **Figure 15** is a schematic representation of the construction of pLOI2352 for the functional integration of *celY* and *celZ* into the chromosome of ethanologenic *K. oxytoca* P2. Coding regions for *celY* and *celZ* are shown as solid segments. Fragments of *Z. mobilis* DNA that serve as promoters (prom1 and prom2) are shown as open segments.

TOP SECRET//SI

Replicons and antibiotic resistance genes are stippled; other vector DNA is shown as a thin connecting line. Arrows on segments indicate the direction of transcription. The small open arrows represent the FRT sites, which are recognized by the *flp* recombinase. FRT sequences are asymmetrical and arranged to allow the deletion of plasmid DNA

5 (replicon and selectable marker) after chromosomal integration.

**Figure 16** is a schematic representation of the construction of pLOI2357 for the inactivation of *celZ* by double homologous recombination. Coding regions for *celY* are shown as solid segments. The fragment of *Z. mobilis* DNA that serves as a promoter (prom1) is shown as open segments. Replicons and antibiotic resistance genes are

10 stippled; other vector DNA is shown as a thin connecting line. Arrows on segments indicate the direction of transcription. The small open arrows represent the FRT sites, which are recognized by the *flp* recombinase.

**Figure 17** shows a digital image of a thin layer chromatogram of fermentation broth illustrating the utilization of cellobiosides. The left panel (A) represents strain

15 P2(pCPP2006), the parent lacking *E. chrysanthemi* endoglucanases, whereas the right panel (B) represents strain SZ21(pCPP2006), the recombinant secreting high levels of CelY and CelZ endoglucanases. Approximately 4  $\mu$ L of broth was spotted in each lane and labels G1 through G6 refer to the number of glucosyl residues in the adjacent spots.

20 Lanes for both panels (A and B) are from left to right: 1, initial (0 h); 2, after 10 h; and 3, after 36 h of fermentation.

**Figure 18** shows a graphical depiction of the amount of ethanol production from amorphous cellulose by ethanologenic derivatives of *K. oxytoca* M5A1. Results shown in the top and middle panel (A and B) include standard deviations from three replicates; results shown in the bottom panel (C) represent single fermentations. The top panel (A) shows results from the fermentation of 6.85 g/L amorphous cellulose; the middle panel (B) shows results from the fermentation of 15.33 g/L amorphous cellulose; and the bottom panel (C) shows the fermentation of 28.96 g/L amorphous cellulose. Strain P2(pCPP2006) is the parental strain and lacks *E. chrysanthemi* endoglucanase genes. Strain SZ6(pCPP2006) secretes CelZ endoglucanase; strain SZ21(pCPP2006) secretes

25 CelY and CelZ endoglucanases; strain SZ22(pCPP2006) secretes CelY endoglucanase.

30

### **Detailed Description of the Invention**

In order for the full scope of the invention to be clearly understood, the following definitions are provided.

5

#### ***I. Definitions***

As used herein the term "recombinant host" is intended to include a cell suitable for genetic manipulation, *e.g.*, which can incorporate heterologous polynucleotide sequences, *e.g.*, which can be transfected. The cell can be a microorganism or a higher 10 eukaryotic cell. The term is intended to include progeny of the cell originally transfected. In preferred embodiments, the cell is a bacterial cell, *e.g.*, a Gram-negative bacterial cell, and this term is intended to include all facultatively anaerobic Gram-negative cells of the family Enterobacteriaceae such as *Escherichia*, *Shigella*, *Citrobacter*, *Salmonella*, *Klebsiella*, *Enterobacter*, *Erwinia*, *Kluyvera*, *Serratia*, 15 *Cedecea*, *Morganella*, *Hafnia*, *Edwardsiella*, *Providencia*, *Proteus*, and *Yersinia*. Particularly preferred recombinant hosts are *Escherichia coli* or *Klebsiella oxytoca* cells.

The term "ratio" is intended to include the relationship between the amounts (measured, *e.g.*, by activity or in moles) of two enzymes in a predetermined combination where, preferably, the ratio is not naturally occurring and, more preferably, results in 20 synergistic enzyme activity.

The terms "a first endoglucanase having a first degrading activity" and "a second endoglucanase having a second degrading activity" are intended to include, respectively, an endoglucanase with an activity that can be distinguished from another endoglucanase (*e.g.*, a second endoglucanase) with a second activity functionally (*e.g.*, by its activity on 25 a particular substrate; synergism with another enzyme), by source of origin (*e.g.*, host cell strain, including naturally occurring strains or genetically modified strains expressing a clone expressing an endoglucanase), or by biochemical properties using art recognized techniques (*e.g.*, molecular weight determination or purification characteristics). The degrading activity, *e.g.*, the enzymatic hydrolysis of an 30 oligosaccharide, can also comprise a change in the viscosity of the oligosaccharide.

The terms "synergism," "synergistic activity," and "synergized" are intended to describe the interaction between distinguishable polypeptides or polypeptide activities wherein the effect of the total activity of the polypeptides taken together are greater than the sum of the effects of the individual activities. The polypeptides may be, for 35 example, endoglucanases, exoglucanases, cellobiohydrolases,  $\beta$ -glucosidases, endo-1,4- $\beta$ -xylanases,  $\alpha$ -xylosidases,  $\alpha$ -glucuronidases,  $\alpha$ -L-arabinofuranosidases, acetylerestases, acetylxylanesterases,  $\alpha$ -amylases,  $\beta$ -amylases, glucoamylases,

50 55 60 65 70 75 80 85 90 95 100

- pullulanases,  $\beta$ -glucanases, hemicellulases, arabinosidases, mannanases, pectin hydrolases, pectate lyases, or any combination thereof. An activity of a polypeptide includes the degradation (*e.g.*, hydrolysis) of an oligosaccharide but may also include a change in the viscosity of the oligosaccharide. The synergized degrading of an
- 5      oligosaccharide is preferably by a factor of about 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, or more preferably about 2.0, with a typical factor being about 1.8.

The term "heterologous polynucleotide segment" is intended to include a polynucleotide segment that encodes one or more polypeptides or portions or fragments of polypeptides. A heterologous polynucleotide segment may be derived from any

10     source, *e.g.*, eukaryotes, prokaryotes, virii, or synthetic polynucleotide fragments.

The terms "polysaccharase," "cellulase," or "glucanase" are used interchangeably herein and are intended to include a polypeptide capable of catalyzing the degradation or depolymerization of any linked sugar moiety, *e.g.*, disaccharides, trisaccharides, oligosaccharides, including, complex carbohydrates, also referred to

15     herein as complex sugars, *e.g.*, cellobiosaccharide and lignocellulose, which comprises cellulose, hemicellulose, and pectin. The terms are intended to include cellulases such as glucanases, including preferably, endoglucanases but also including, *e.g.*, exoglucanase,  $\beta$ -glucosidase, cellobiohydrolase, endo-1,4- $\beta$ -xyylanase,  $\beta$ -xylosidase,  $\alpha$ -glucuronidase,  $\alpha$ -L-arabinofuranosidase, acetyl esterase, acetyl xylan esterase,  $\alpha$ -amylase,

20      $\beta$ -amylase, glucoamylase, pullulanase,  $\beta$ -glucanase, hemicellulase, arabinosidase, mannanase, pectin hydrolase, pectate lyase, or a combination of any of these cellulases.

The term "endoglucanase" is intended to include a cellulase which typically hydrolyses internal  $\beta$ 1-4 glucosyl linkages in polymeric substrates and does not preferentially hydrolyze linkages located at the ends of the chain.

25     The term "surrogate promoter" is intended to include a polynucleotide segment that can transcriptionally control a gene-of-interest that it does not transcriptionally control in nature. In a preferred embodiment, the transcriptional control of a surrogate promoter results in an increase in expression of the gene-of-interest. In a preferred embodiment, a surrogate promoter is placed 5' to the gene-of-interest. A surrogate

30     promoter may be used to replace the natural promoter, or may be used in addition to the natural promoter. A surrogate promoter may be endogenous with regard to the host cell in which it is used or it may be a heterologous polynucleotide sequence introduced into the host cell, *e.g.*, exogenous with regard to the host cell in which it is used. Other promoters suitable for use in bacteria include, *e.g.*, lacZ, T7, and SP6 (see, *e.g.*, Ausubel et al. *infra*).

The terms "oligosaccharide source," "oligosaccharide," "complex cellulose," "complex carbohydrate," and "complex sugar," and "polysaccharide" are used

- essentially interchangeably and are intended to include any carbohydrate source comprising more than one sugar molecule. These carbohydrates may be derived from any unprocessed plant material or any processed plant material. Examples are wood, paper, pulp, plant derived fiber, or synthetic fiber comprising more than one linked
- 5 carbohydrate moiety, *i.e.*, one sugar residue. One particular oligosaccharide source is lignocellulose, which represents approximately 90% of the dry weight of most plant material and contains carbohydrates, *e.g.*, cellulose, hemicellulose, pectin, and aromatic polymers, *e.g.*, lignin. Cellulose makes up 30%-50% of the dry weight of lignocellulose and is a homopolymer of cellobiose (a dimer of glucose). Similarly, hemicellulose,
- 10 makes up 20%-50% of the dry weight of lignocellulose and is a complex polymer containing a mixture of pentose (xylose, arabinose) and hexose (glucose, mannose, galactose) sugars which contain acetyl and glucuronyl side chains. Pectin makes up 1%-20% of the dry weight of lignocellulose and is a methylated homopolymer of glucuronic acid. Other oligosaccharide sources include carboxymethyl cellulose (CMC),
- 15 amorphous cellulose (*e.g.*, acid-swollen cellulose), and the cellooligosaccharides cellobiose, cellotriose, cellotetraose, and cellopentaose. Cellulose, *e.g.*, amorphous cellulose may be derived from a paper or pulp source (including, *e.g.*, fluid wastes thereof) or, *e.g.*, agricultural byproducts, *e.g.*, corn stalks, soybean solubles, or beet pulp. Any one or a combination of the above carbohydrate polymers are potential sources of
- 20 sugars for depolymerization and subsequent bioconversion to ethanol by fermentation according to the products and methods of the present invention.

The term "gene/s" or "polynucleotide segment" is intended to include nucleic acid molecules, *e.g.*, polynucleotides which include an open reading frame encoding a polypeptide, and can further include non-coding regulatory sequences, and introns. In

25 addition, the terms are intended to include one or more genes that map to a functional locus, *e.g.*, the *out* or *pul* genes of *Erwinia* and *Klebsiella*, respectively, that encode more than one gene product, *e.g.*, a secretory polypeptide. In addition, the terms are intended to include a specific gene for a selected purpose. The gene may be endogenous to the host cell or may be recombinantly introduced into the host cell, *e.g.*, as a plasmid

30 maintained episomally or a plasmid (or fragment thereof) that is stably integrated into the genome. In a preferred embodiment, the gene of polynucleotide segment is involved in at least one step in the bioconversion of a carbohydrate to ethanol. Accordingly, the term is intended to include any gene encoding a polypeptide such as an alcohol dehydrogenase, a pyruvate decarboxylase, a secretory protein/s, or a polysaccharase,

35 *e.g.*, a glucanase, such as an endoglucanase or exoglucanase, a cellobiohydrolase,  $\beta$ -glucosidase, endo-1,4- $\beta$ -xylanase,  $\beta$ -xylosidase,  $\alpha$ -glucuronidase,  $\alpha$ -L-arabinofuranosidase, acetylersterase, acetylxylanesterase,  $\alpha$ -amylase,  $\beta$ -amylase,

glucoamylase, pullulanase,  $\beta$ -glucanase, hemicellulase, arabinosidase, mannanase, pectin hydrolase, pectate lyase, or a combination thereof.

The term "simultaneous saccharification and fermentation" or "SSF" is intended to include the use of one or more recombinant hosts (or extracts thereof, including

- 5 purified or unpurified extracts) for the contemporaneous degradation or depolymerization of a complex sugar and bioconversion of that sugar residue into ethanol by fermentation.

The term "transcriptional control" is intended to include the ability to modulate gene expression at the level of transcription. In a preferred embodiment, transcription,

- 10 and thus gene expression, is modulated by replacing or adding a surrogate promoter near the 5' end of the coding region of a gene-of-interest thereby resulting in altered gene expression. In a most preferred embodiment, the transcriptional control of one or more gene is engineered to result in the optimal expression of such genes, *e.g.*, in a desired ratio. The term also includes inducible transcriptional control as recognized in the art.

- 15 The term "expression" is intended to include the expression of a gene at least at the level of mRNA production.

The term "expression product" is intended to include the resultant product of an expressed gene, *e.g.*, a polypeptide.

- 20 The term "increased expression" is intended to include an alteration in gene expression at least at the level of increased mRNA production and preferably, at the level of polypeptide expression.

The term "increased production" is intended to include an increase in the amount of a polypeptide expressed, in the level of the enzymatic activity of the polypeptide, or a combination thereof.

- 25 The terms "activity" and "enzymatic activity" are used interchangeably and are intended to include any functional activity normally attributed to a selected polypeptide when produced under favorable conditions. The activity of an endoglucanase (*e.g.*, EGY or EGZ) is, for example, the ability of the polypeptide to enzymatically depolymerize a complex saccharide. Typically, the activity of a selected polypeptide

- 30 encompasses the total enzymatic activity associated with the produced polypeptide. The polypeptide produced by a host cell and having enzymatic activity may be located in the intracellular space of the cell, cell-associated, secreted into the extracellular milieu, or a combination thereof. Techniques for determining total activity as compared to secreted activity are described herein and are known in the art.

- 35 The term "secreted" is intended to include an increase in the secretion of a polypeptide into the periplasmic space or into the extracellular milieu, *e.g.*, a heterologous polypeptide, preferably a polysaccharase. Typically, the polypeptide is

secreted at an increased level that is in excess of the naturally-occurring amount of secretion. More preferably, the term "secreted" refers to an increase in secretion of a given polypeptide that is at least 10% and more preferably, at least about 100%, 200%, 300%, 400%, 500%, 600%, 700%, 800%, 900%, 1000%, or more, as compared to the naturally-occurring level of secretion.

The term "secretory polypeptide" is intended to include any polypeptide/s, alone or in combination with other polypeptides, that facilitate the transport of another polypeptide from the intracellular space of a cell to the extracellular milieu. In one embodiment, the secretory polypeptide/s encompass all the necessary secretory polypeptides sufficient to impart secretory activity to a Gram-negative host cell. Typically, secretory proteins are encoded in a single region or locus that may be isolated from one host cell and transferred to another host cell using genetic engineering. In a preferred embodiment, the secretory polypeptide/s are derived from any bacterial cell having secretory activity. In a more preferred embodiment, the secretory polypeptide/s are derived from a host cell having Type II secretory activity. In another more preferred embodiment, the host cell is selected from the family Enterobacteriaceae. In a most preferred embodiment, the secretory polypeptide/s are one or more gene products of the *out* or *pul* genes derived from, respectively, *Erwinia* or *Klebsiella*. Moreover, the skilled artisan will appreciate that any secretory protein/s derived from a related host that is sufficiently homologous to the *out* or *pul* gene/s described herein may also be employed (Pugsley *et al.*, (1993) *Microbiological Reviews* 57:50-108; Lindeberg *et al.*, (1996) *Mol. Micro.* 20:175-190; Lindeberg *et al.*, (1992) *J. of Bacteriology* 174:7385-7397; He *et al.*, (1991) *Proc. Natl. Acad. Sci. USA*, 88:1079-1083).

The term "derived from" is intended to include the isolation (in whole or in part) of a polynucleotide segment from an indicated source or the purification of a polypeptide from an indicated source. The term is intended to include, for example, direct cloning, PCR amplification, or artificial synthesis from, or based on, a sequence associated with the indicated polynucleotide source.

The term "ethanogenic" is intended to include the ability of a microorganism to produce ethanol from a carbohydrate as a primary fermentation product. The term includes but is not limited to naturally occurring ethanogenic organisms, organisms with naturally occurring or induced mutations, and organisms that have been genetically modified.

The term "Gram-negative bacteria" is intended to include the art recognized definition of this term. Typically, Gram-negative bacteria include, for example, the family Enterobacteriaceae which comprises, among others, the species *Escherichia* and *Klebsiella*.

- The term "sufficiently homologous" is intended to include a first amino acid or nucleotide sequence which contains a sufficient or minimum number of identical or equivalent amino acid residues or nucleotides, *e.g.*, an amino acid residue which has a similar side chain, to a second amino acid or nucleotide sequence such that the first and
- 5 second amino acid or nucleotide sequences share common structural domains and/or a common functional activity. For example, amino acid or nucleotide sequences which share common structural domains have at least about 40% homology, preferably 50% homology, more preferably 60%, 70%, 80%, or 90% homology across the amino acid sequences of the domains and contain at least one, preferably two, more preferably
- 10 three, and even more preferably four, five, or six structural domains, are defined herein as sufficiently homologous. Furthermore, amino acid or nucleotide sequences which share at least 40%, preferably 50%, more preferably 60%, 70%, 80%, or 90% homology and share a common functional activity are defined herein as sufficiently homologous.

In one embodiment, two polynucleotide segments, *e.g.*, promoters, are

15 "sufficiently homologous" if they have substantially the same regulatory effect as a result of a substantial identity in nucleotide sequence. Typically, "sufficiently homologous" sequences are at least 50%, more preferably at least 60%, 70%, 80%, or 90% identical, at least in regions known to be involved in the desired regulation. More preferably, no more than five bases differ. Most preferably, no more than five

20 consecutive bases differ.

To determine the percent identity of two polynucleotide segments, or two amino acid sequences, the sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for

25 comparison purposes). In a preferred embodiment, the length of a reference sequence aligned for comparison purposes is at least 30%, preferably at least 40%, more preferably at least 50%, even more preferably at least 60%, and even more preferably at least 70%, 80%, or 90% of the length of the reference sequence. The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are

30 then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position (as used herein amino acid or nucleic acid "identity" is equivalent to amino acid or nucleic acid "homology"). The percent identity between the two sequences is a function of the number of identical positions

35 shared by the sequences, taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences.

The comparison of sequences and determination of percent identity between two sequences can be accomplished using a mathematical algorithm. In a preferred embodiment, the percent identity between two amino acid sequences is determined using the Needleman and Wunsch (*J. Mol. Biol.* (48):444-453 (1970)) algorithm which

5 has been incorporated into the GAP program in the GCG software package (available at <http://www.gcg.com>), using either a Blossom 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. In yet another preferred embodiment, the percent identity between two nucleotide sequences is determined using the GAP program in the GCG software package (available at

10 <http://www.gcg.com>), using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. In another embodiment, the percent identity between two amino acid or nucleotide sequences is determined using the algorithm of E. Meyers and W. Miller (CABIOS, 4:11-17 (1989)) which has been incorporated into the ALIGN program (version 2.0), using a PAM120 weight residue

15 table, a gap length penalty of 12 and a gap penalty of 4.

The polynucleotide and amino acid sequences of the present invention can further be used as a "query sequence" to perform a search against public databases to, for example, identify other family members or related sequences, e.g., promoter sequences. Such searches can be performed using the NBLAST and XBLAST programs

20 (version 2.0) of Altschul, *et al.* (1990) *J. Mol. Biol.* 215:403-10. BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to polynucleotide molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to polypeptide molecules of

25 the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul *et al.*, (1997) *Nucleic Acids Res.* 25(17):3389-3402. When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used. See <http://www.ncbi.nlm.nih.gov>.

30

## ***II. Synergism Between Endoglucanases***

The present invention is based, at least in part, on the discovery that endoglucanases act synergistically in degrading complex sugars. This invention is based, also in part, on the functional integration and expression of two endoglucanases

35 (e.g., EGY and EGZ) by an ethanogenic host cell (e.g., *K. oxytoca* P2) to effect synergistic degradation of oligosaccharides (e.g., crystalline cellulose) and increase the production of ethanol by simultaneous saccharification and fermentation (SSF).

In one embodiment, an endoglucanase is derived from *E. chrysanthemi* and is the endoglucanase EGZ, which is encoded by the *celZ* gene (Boyer, *et al.* (1987) *Eur. J. Biochem.* 162:311-316). In another embodiment, an endoglucanase is derived from *E. chrysanthemi* and is the endoglucanase EGY, which is encoded by the *celY* gene

5 (Guiseppi *et al.*, (1991) *Gene* 106:109-114). *E. chrysanthemi* EGY and EGZ are endoglucanases that have high activities in the degradation of carboxymethyl cellulose (CMC) and belong to, respectively, Type IV and Type II secretion groups (Hueck *et al.* (1998) *Micro and Mol Biol Rev* 62:379-433). EGY and EGZ differ in substrate range and function synergistically during the hydrolysis of CMC and amorphous cellulose,

10 indicating a potential need for both enzymes for optimal cellulase activity. Specifically, EGZ hydrolyzes cellotriose, cellotetraose, cellopentaose, amorphous cellulose, and CMC. EGY hydrolyses polymeric substrates to products of approximately 10 glucosyl residues.

In another embodiment, the endoglucanases (*e.g.*, EGY and EGZ) are purified

15 separately and combined in a ratio sufficient for the synergistic degradation of an oligosaccharide substrate to occur and this may be determined using the assays disclosed herein. These assays allow for a determination and optimization of a ratio between, *e.g.*, two given glucanases, *e.g.*, endoglucanases. Typically the ratios range from about 9 to 1 to about 19 to 1. In one embodiment, the ratio can be about 9 to 1 or 19 to 1 for EGZ to

20 EGY. In a preferred embodiment, optimum synergy is observed with a high ratio of EGZ to EGY, similar to that produced by *E. chrysanthemi* and by SZ21, a *K. oxytoca* recombinant which expresses *celY* and *celZ* (See example 4).

In yet another embodiment, the endoglucanases can be combined concurrently with the oligosaccharide substrate. In yet another embodiment, the endoglucanases can

25 be added to the oligosaccharide substrate sequentially. In a preferred embodiment, EGZ is sequentially added to the substrate following the addition of EGY, after heat inactivation of EGY activity. Example 3 describes, in detail, the synergistic effect of various ratios of endoglucanases (*e.g.*, EGY and EGZ), the synergistic effect of sequential addition of endoglucanases (*e.g.*, EGY and EGZ) (See Table 12), and the

30 effect of various substrate concentrations (See Table 11) and incubation times on the synergistic activity of the endoglucanases.

In yet another embodiment, the synergistic degradation of the oligosaccharide is of a factor of about 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, or 2.0 with a more preferable factor being about 1.8. The synergistic factor is calculated as the observed degradation

35 divided by the sum of predicted contributions from EGY alone and EGZ alone (Riedel, *et al.* (1997) *FEMS Microbiol. Lett.* 147:239-243). In one embodiment, the endoglucanases are EGY and EGZ. The predicted contribution of EGY alone is about

10% of the total activity, and the predicted contributions of EGZ alone are about 90% of the total activity.

In another aspect of the invention, at least one of the endoglucanases is derived from a cell extract. The cell may be recombinantly engineered to produce at least one endoglucanase. In a preferred embodiment, the cell is recombinantly engineered to produce two endoglucanases (e.g., EGY and EGZ). For example, the cell can be a bacterial cell. The recombinant cell comprises at least one heterologous polynucleotide segment, or, preferably, two or more heterologous polynucleotides segments, encoding a polypeptide/s under the transcriptional control of one or more heterologous surrogate promoter/s. The heterologous polynucleotide and surrogate promoter may be plasmid based or integrated into the genome of the organism (as described in the examples). In a preferred embodiment, the host cell is used as a source of a desired polypeptide for use in the bioconversion of a complex sugar to ethanol, or a step thereof. In another preferred embodiment, the heterologous polynucleotide segment encodes one or more endoglucanases (e.g., EGY or EGZ) which are expressed at higher levels than are naturally occurring in the host. In one embodiment, the endoglucanases are purified separately from different recombinant cells and subsequently combined to synergistically degrade a substrate (e.g., an oligosaccharide). In another embodiment, one recombinant host cell can produce two or more endoglucanases concurrently and can act synergistically to degrade a substrate.

In another aspect of the invention, the recombinant bacterial host cell is an ethanogenic bacterium such as, for example, *K. oxytoca* P2, an ethanologic derivative of M5A1 (Wood, *et al.* (1992) *Appl. Environ. Microbiol.* 58:2103-2110). In one embodiment, the recombinant ethanogenic bacterium contains at least one heterologous polynucleotide segment (e.g., *celY* or *celZ* derived from *Erwinia*) encoding at least one endoglucanase (e.g., EGY or EGZ). In a preferred embodiment, the recombinant ethanogenic bacteria contains more than one heterologous polynucleotide segments which encode endoglucanases. For example, as described in detail in Example 4, *celY* and *celZ* can be functionally integrated, expressed, and secreted from the ethanologic strain *K. oxytoca* P2 concurrently to produce ethanol from an oligosaccharide substrate (e.g., crystalline cellulose).

*Sub B4* In another embodiment, the recombinant host is a Gram-negative bacterium. In yet another embodiment, the recombinant host is from the family Enterobacteriaceae. The ethanogenic hosts of U.S.P.N. 5,821,093, hereby incorporated by reference, for example, are suitable hosts and include, in particular, *E. coli* strains KO4 (ATCC 55123), KO11 (ATCC 55124), and KO12 (ATCC 55125), and *Klebsiella oxytoca* strain P2 (ATCC 55307). Alternatively, a non-ethanogenic host of the present invention

may be converted into an ethanologenic host (such as the above-mentioned strains) by introducing, for example, ethanologenic genes from an efficient ethanol producer like *Zymomonas mobilis*. This type of genetic engineering, using standard techniques, results in a recombinant host capable of efficiently fermenting sugar into ethanol. In 5 addition, the LY01 ethanol tolerant strain may be employed as described in published PCT international application WO 98/45425 and this published application is hereby incorporated by reference (see also, e.g., Yomano *et al.* (1998) *J. of Ind. Micro. & Bio.* 20:132-138).

In another preferred embodiment, the invention makes use of a non-  
10 ethanologenic recombinant host, e.g., *E. coli* strain B, *E. coli* strain DH5 $\alpha$ , or *Klebsiella oxytoca* strain M5A1. These strains may be used to express at least one desired polypeptide, e.g., an endoglucanase, using techniques described herein. In addition, these recombinant hosts may be used in conjunction with another recombinant host that expresses yet another desirable polypeptide, e.g., a different endoglucanase. For  
15 example, a recombinant host producing EGZ can be combined with a recombinant host producing EGY to produce a synergistic effect. In addition, the non-ethanologenic host cell/s may be used in conjunction with an ethanologenic host cell. For example, the use of a non-ethanologenic host/s for carrying out, e.g., the synergistic depolymerization of a complex sugar may be followed by the use of an ethanologenic host for fermenting the  
20 depolymerized sugar. Accordingly, it will be appreciated that these reactions may be carried out serially or contemporaneously using, e.g., homogeneous or mixed cultures of non-ethanologenic and ethanologenic recombinant hosts.

In a preferred embodiment, one or more genes for fermenting a sugar substrate into ethanol are provided on a plasmid or integrated into the host chromosome. More  
25 preferably, genes for fermenting a sugar substrate into ethanol, e.g., pyruvate decarboxylase (e.g., *pdc*) and/or alcohol dehydrogenase (e.g., *adh*) are introduced into the host of the invention using an artificial operon such as the PET operon as described in U.S.P.N. 5,821,093, hereby incorporated by reference. Indeed, it will be appreciated that the present invention, in combination with what is known in the art, provides  
30 techniques and vectors for introducing multiple genes into a suitable host (see, e.g., *Current Protocols in Molecular Biology*, eds. Ausubel *et al.*, John Wiley & Sons (1992), Sambrook, J. *et al.*, *Molecular Cloning: A Laboratory Manual*. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY (1989), and *Bergey's Manual of Determinative Bacteriology*, Kreig *et al.*, Williams and  
35 Wilkins (1984), hereby incorporated by reference).

Accordingly, using the methods of the invention, a single genetic construct can encode all of the necessary gene products (e.g., a glucanase, an endoglucanase, an

- exoglucanase, a secretory protein/s, pyruvate decarboxylase, alcohol dehydrogenase) for performing simultaneous saccharification and fermentation (SSF). For example, Example 4 describes, in detail, the simultaneous saccharification and fermentation (SSF) of crystalline cellulose (Sigmacell 50) by bacterial cellulases EGY and EGZ produced
- 5 by ethanologenic *K. oxytoca*, with added commercial cellulase (Spezyme®). The endoglucanases produced by ethanologenic *K. oxytoca* and the commercial cellulase (Spezyme®) function synergistically to increase ethanol production (7% to 22%) from crystalline cellulose (Sigmacell 50). The beneficial effect is attributed almost exclusively to EGY, despite the fact that EGY activities were low in comparison to
- 10 EGZ. Activity of the ethanologenic *K. oxytoca* strain SZ22, which expresses EGY, was nearly equivalent to the activity of the ethanologenic *K. oxytoca* strain SZ21, which expresses both EGY and EGZ activities. *K. oxytoca* strain SZ6, which expresses only EGZ showed little benefit from the production of over 20,000 U of endoglucanase activity per liter.
- 15 In one embodiment, the composition of two endoglucanases which act synergistically to degrade an oligosaccharide also includes at least one additional enzymatic activity. This additional activity may be a glucanase activity selected from the group consisting of endoglucanase, exoglucanase, cellobiohydrolase,  $\beta$ -glucosidase, endo-1,4- $\beta$ -xylanase,  $\alpha$ -xylosidase,  $\alpha$ -glucuronidase,  $\alpha$ -L-arabinofuranosidase,
- 20 acetylesterase, acetylxylanesterase,  $\alpha$ -amylase,  $\beta$ -amylase, glucoamylase, pullulanase,  $\beta$ -glucanase, hemicellulase, arabinosidase, mannanase, pectin hydrolase, pectate lyase, or a combination thereof. In another embodiment, this additional enzymatic activity may be derived from a fungus, for example *T. longibrachiatum*. Fungi such as *T. longibrachiatum* produce multiple endoglucanase activities, which are presumed to
- 25 function together with exoglucanases during the hydrolysis of crystalline cellulose (Nidetzky, et al. (1995) *Synergistic interaction of cellulases from Trichoderma reesei during cellulose degradation* p.90-112. In J. N. Saddler and M. E. Himmel (ed.). *Enzymatic degradation of insoluble carbohydrates* ACS symposium series 618, American Chemical Society, Washington, D.C.; Tomme, et al. (1995) *Adv. Microbiol. Physiol.* 37:1-81; Woodward, J. (1991) *Bioresource Technol.* 36:67-75).
- 30 In contrast to EGZ, EGY does not hydrolyze soluble cellobiosides but preferentially acts on longer chain substrates, producing ends, which can function as new sites for exoglucanase activity. In the absence of fungal cellulase additions, EGY and EGZ function synergistically to degrade amorphous cellulose. In nature,
- 35 lignocellulosic substrates are depolymerized by mixtures of extracellular enzymes produced by consortia of fungi and bacteria. Thus, a mixture of *E. chrysanthemi*

enzymes and enzymes from the fungus *T. reesei* can improve the digestion of lignocellulosic substrates during bioconversion to ethanol.

It will also be appreciated that a recombinant host may be further manipulated, using methods known in the art, to have mutations in any endogenous gene/s (*e.g.*, recombinase genes) that would interfere with the stability, expression, function, and secretion of the introduced genes. Further, it will also be appreciated that the invention is intended to encompass any regulatory elements, gene/s, or gene products, *i.e.*, polypeptides, that are sufficiently homologous to the ones described herein.

For effective degradation of oligosaccharides, the glucanase (*e.g.*, EGY or EGZ) is preferably secreted into the extracellular milieu. Accordingly, in another embodiment of the invention, the host cell has been engineered to express a secretory protein/s to facilitate the export of the desired polypeptide from the cell. In one embodiment, the secretory protein or proteins are derived from a Gram-negative bacterial cell, *e.g.*, a cell from the family Enterobacteriaceae. In another embodiment, the secretory protein/s are from *Erwinia* and are encoded by the *out* genes. In another embodiment, the secretory proteins are the *pul* genes derived from *Klebsiella*. The introduction of one or more of these secretory proteins is especially desirable if the host cell is an enteric bacterium, *e.g.*, a Gram-negative bacterium having a cell wall. Representative Gram-negative host cells of the invention are from the family Enterobacteriaceae and include, *e.g.*, *Escherichia* and *Klebsiella*. In one embodiment, the introduction of one or more secretory proteins into the host results in an increase in the secretion of the selected protein, *e.g.*, a glucanase, as compared to naturally-occurring levels of secretion. Preferably, the increase in secretion is at least about 10% and more preferably, 100%, 200%, 300%, 400%, 500%, 600%, 700%, 800%, 900%, 1000%, or more, as compared to naturally-occurring levels of secretion. In a preferred embodiment, the addition of secretion genes allows for the glucanase polypeptide to be produced at higher levels. In a preferred embodiment, the addition of secretion genes allows for the glucanase polypeptide to be produced with higher enzymatic activity. In a most preferred embodiment, the glucanase is produced at higher levels and with higher enzymatic activity. Preferably, an increase in glucanase activity of at least about 10%, more preferably about 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, or 100% is observed. Most preferably, an increase in glucanase activity of several fold is obtained, *e.g.*, about 200%, 300%, 400%, 500%, 600%, 700%, 800%, 900%, or 1000%, as compared to cells without secretion genes (*e.g.*, cells that either lack or do not express secretion genes at a sufficient level). The techniques and methods for introducing such genes and measuring increased output of a desired polypeptide such as, *e.g.*, a

glucanase, are described in further detail in the examples. Other equivalent methods are known to those skilled in the art.

EGY and EGZ are secreted by different mechanisms. As described in Example 4, approximately 70% of the EGZ produced was secreted as an extracellular product 5 when the *E. chrysanthemi* *out* genes were added on a plasmid (pCPP2006) consistent with a Type II secretion system (Hueck, C.J. (1998) *Microbiol. Mol. Biol. Rev.* 62:379-433). Half of the EGY activity was secreted in the presence or absence of the *out* genes consistent with a Type IV secretion system (Hueck, C.J. *supra*).

Methods for screening strains having the introduced genes (*e.g.*, endoglucanase 10 encoding genes such as *celY* and *celZ* and alcohol dehydrogenase genes) are routine and may be facilitated by visual screens that can identify cells expressing either the alcohol dehydrogenase (ADH) or glucanase (*e.g.*, EGZ or EGY) gene product. The ADH gene product produces acetaldehyde that reacts with the leucosulfonic acid derivative of p-roseaniline to produce an intensely red product. Thus, ADH-positive clones can be 15 easily screened and identified as bleeding red colonies. Methods for screening for a polysaccharase activity (*e.g.*, an endoglucanase , such as EGZ or EGY), also results in a clear visual phenotype as described below and in the examples.

Methods for screening for synergism between two endoglucanases are described 20 in detail in the examples. EGY and EGZ can be purified separately from recombinant host cells transformed with a plasmid containing the *celY* gene or the *celZ* gene. The cell-free culture broth can be used to determine extracellular endoglucanase activity. Broth containing cells disrupted by ultrasound can be used to determine total activity. Hydrolysis of celooligosaccharides can be analyzed by thin layer chromatography. Additionally, endoglucanase activity using CMC as a substrate can be determined *in* 25 *vitro* by analyzing samples for reducing sugars. Reducing sugars can be measured using 3,5-dinitrosalicylic acid reagent with glucose as a standard. Synergy is calculated as the observed activity divided by the sum of predicted contributions from EGY alone (10%) plus EGZ alone (90%).

Alternatively, recombinant host cells can be transformed with a plasmid 30 containing both endoglucanases to produce both endoglucanases concurrently. Synergy can be determined as described above. Simultaneous saccharification and fermentation (SSF) can be carried out to observe ethanol production from a substrate (as described in Example 4). In SSF experiments with added commercial cellulase, two of the *K. oxytoca* recombinants containing at least one *E. chrysanthemi* endoglucanase produced 35 more ethanol than the parent *K. oxytoca* strain which lacked the *E. chrysanthemi* endoglucanases. Both of these strains also produced ethanol levels equivalent to the best yeast SSF experiments (Cho, *et al.* (1999) *J. Microbiol. Biotechnol.* 9:340-345)

using approximately one-third of the amount of added commercial cellulase (500 FPU/g cellulose versus 18 FPU/g cellulose for recombinant yeast).

Recombinant bacteria expressing, for example, the PET operon typically grow to higher cell densities in liquid culture than the unmodified parent organisms due to the production of neutral rather than acidic fermentation products (Ingram *et al.*, (1988) *Appl. Environ. Microbiol.* 54:397-404). On plates, ethanologenic clones are readily apparent as large, raised colonies which appear much like yeast. These traits have been very useful during the construction of new strains and can provide a preliminary indication of the utility of new constructs. Rapid evaluations of ethanol producing potential can also be made by testing the speed of red spot development on aldehyde indicator plates (Conway *et al.*, (1987) *J. Bacteriol.* 169:2591-2597). Typically, strains which prove to be efficient in sugar conversion to ethanol can be recognized by the production of red spots on aldehyde indicator plates within minutes of transfer.

In a most preferred embodiment of the invention, a single host cell is ethanologenic, that is, has all the necessary genes, either naturally occurring or artificially introduced or enhanced (*e.g.*, using a surrogate promoter and/or genes from a different species or strain), such that the host cell has the ability to produce and secrete two glucanses, preferably, an endoglucanase, more preferably at least two endoglucanases sufficient to , degrade a complex sugar, and ferment the degraded sugar into ethanol. Accordingly, such a host is suitable for simultaneous saccharification and fermentation.

Moreover, the present invention takes into account that the native *E. coli* fermentation pathways produce a mixture of acidic and neutral products (in order of abundance): lactic acid, hydrogen + carbon dioxide (from formate), acetic acid, ethanol, and succinate. However, the *Z. mobilis* PDC (pyruvate decarboxylase) has a lower Km for pyruvate than any of the competing *E. coli* enzymes. By expressing high activities of PDC, carbon flow is effectively redirected from lactic acid and acetyl-CoA into acetylaldehyde and ethanol. Small amounts of succinate can be eliminated by deleting the fumarate reductase gene (*frd*) (Ingram *et al.*, (1991) U.S.P.N, 5,000,000; Ohta *et al.*, (1991) *Appl. Environ. Microbiol.* 57:893-900). Additional mutations (*e.g.*, in the *pfl* or *ldh* genes) may be made to completely eliminate other competing pathways (Ingram *et al.*, (1991) U.S.P.N, 5,000,000). Additional mutations to remove enzymes (*e.g.*, recombinases, such as *recA*) that may compromise the stability of the introduced genes (either plasmid-based or integrated into the genome) may also be introduced, selected for, or chosen from a particular background.

In addition, it should be readily apparent to one skilled in the art that the ability conferred by the present invention, to transform genes coding for a protein or an entire

metabolic pathway into a single manipulable construct, is extremely useful. Envisioned in this regard, for example, is the application of the present invention to a variety of situations where genes from different genetic loci are placed on a chromosome. This may be a multi-cistronic cassette under the control of a single promoter or separate promoters may be used.

- 5 Exemplary *E. coli* strains that are ethanologenic and suitable for further improvement according to the methods of the invention include, for example, KO4, KO11, and KO12 strains, as well as the LY01 strain, an ethanol-tolerant mutant of the *E. coli* strain KO11. Ideally, these strains may be derived from the *E. coli* strain ATCC 10 11303, which is hardy to environmental stresses and can be engineered to be ethanologenic and secrete a polysaccharase/s. In addition, recent PCR investigations have confirmed that the ATCC 11303 strain lacks all genes known to be associated with the pathogenicity of *E. coli* (Kuhnert *et al.*, (1997) *Appl. Environ. Microbiol.* 63:703-709).
- 15 Another preferred ethanologenic host for improvement according to the methods of the invention is the *E. coli* KO11 strain which is capable of fermenting hemicellulose hydrolysates from many different lignocellulosic materials and other substrates (Asghari *et al.*, (1996) *J. Ind. Microbiol.* 16:42-47; Barbosa *et al.*, (1992) *Current Microbiol.* 28:279-282; Beall *et al.*, (1991) *Biotechnol. Bioeng.* 38:296-303; Beall *et al.*, (1992) 20 *Biotechnol. Lett.* 14:857-862; Hahn-Hagerdal *et al.*, (1994) *Appl. Microbiol. Biotechnol.* 41:62-72; Moniruzzaman *et al.*, (1996) *Biotechnol. Lett.* 18:955-990; Moniruzzaman *et al.*, (1998) *Biotechnol. Lett.* 20:943-947; Grohmann *et al.*, (1994) *Biotechnol. Lett.* 16:281-286; Guimaraes *et al.*, (1992) *Biotechnol. Bioeng.* 40:41-45; Guimaraes *et al.*, (1992) *Biotechnol. Lett.* 14:415-420; Moniruzzaman *et al.*, (1997) *J. Bacteriol.* 179:1880-1886). In Figure 1, the kinetics of bioconversion for this strain are shown. In particular, this strain is able to rapidly ferment a hemicellulose hydrolysate from rice hulls (which contained 58.5 g/L of pentose sugars and 37 g/L of hexose sugars) into ethanol (Moniruzzaman *et al.*, (1998) *Biotechnol. Lett.* 20:943-947). It was noted that this strain was capable of fermenting a hemicellulose hydrolysate to completion within 30 48 to 72 hours, and under ideal conditions, within 24 hours.

Another preferred host cell of the invention is the bacterium *Klebsiella*. In particular, *Klebsiella oxytoca* is preferred because, like *E. coli*, this enteric bacterium has the native ability to metabolize monomeric sugars, which are the constituents of more complex sugars. Moreover, *K. oxytoca* has the added advantage of being able to 35 transport and metabolize cellobiose and cellotriose, the soluble intermediates from the enzymatic hydrolysis of cellulose (Lai *et al.*, (1996) *Appl. Environ. Microbiol.* 63:355-363; Moniruzzaman *et al.*, (1997) *Appl. Environ. Microbiol.* 63:4633-4637; Wood *et al.*,

(1992) *Appl. Environ. Microbiol.* 58:2103-2110). The invention provides genetically engineered ethanologenic derivatives of *K. oxytoca*, e.g., strain M5A1 having the *Z. mobilis pdc* and *adhB* genes encoded within the PET operon (as described herein and in U.S.P.N. 5,821,093; Wood *et al.*, (1992) *Appl. Environ. Microbiol.* 58:2103-2110).

- 5        Accordingly, the resulting organism, strain P2, produces ethanol efficiently from monomer sugars and from a variety of saccharides including raffinose, stachyose, sucrose, cellobiose, cellotriose, xylobiose, xylotriose, maltose, etc. (Burchhardt *et al.*, (1992) *Appl. Environ. Microbiol.* 58:1128-1133; Moniruzzaman *et al.*, (1997) *Appl. Environ. Microbiol.* 63:4633-4637; Moniruzzaman *et al.*, (1997) *J. Bacteriol.* 179:1880-10 1886; Wood *et al.*, (1992) *Appl. Environ. Microbiol.* 58:2103-2110). These strains may be further modified according to the methods of the invention to express and secrete one or more polysaccharases (e.g., endoglucanases). Accordingly, this strain is suitable for use in the bioconversion of a complex saccharide in an SSF process (Doran *et al.*, (1993) *Biotechnol. Progress.* 9:533-538; Doran *et al.*, (1994) *Biotechnol. Bioeng.* 44:240-247;
- 15      Wood *et al.*, (1992) *Appl. Environ. Microbiol.* 58:2103-2110). In particular, the use of this ethanologenic P2 strain eliminates the need to add supplemental cellobiase, and this is one of the least stable components of commercial fungal cellulases (Grohmann, (1994) *Biotechnol. Lett.* 16:281-286).
- 20      Screen for Promoters Suitable for Use in Heterologous Gene Expression  
While in one embodiment, the surrogate promoter of the invention is used to improve the expression of a heterologous gene, e.g., a polysaccharase (an endoglucanase for example), it will be appreciated that the invention also allows for the screening of surrogate promoters suitable for enhancing the expression of any desirable gene product.
- 25      In general, the screening method makes use of the cloning vector described in Example 1 and depicted in Figure 3 that allows for candidate promoter fragments to be conveniently ligated and operably-linked to a reporter gene. In one embodiment, the *celZ* gene encoding glucanase serves as a convenient reporter gene because a strong colorimetric change results from the expression of this enzyme (glucanase) when cells
- 30      bearing the plasmid are grown on a particular media (CMC plates). Accordingly, candidate promoters, e.g., a particular promoter sequence or, alternatively, random sequences that can be "shotgun" cloned and operably linked to the vector, can be introduced into a host cell and resultant colonies are scanned, visually, for having increased gene expression as evidenced by a phenotypic glucanase-mediated
- 35      colorimetric change on a CMC plate. Colonies having the desired phenotype are then processed to yield the transforming DNA and the promoter is sequenced using appropriate primers (see Example 1 for more details).

The high correspondence between the glucanase-mediated colorimetric change on a CMC plate and expression levels of the enzyme is an excellent indication of the strength of a candidate promoter (Fig. 4). Hence, the methods of the invention provide a rapid visual test for rating the strength of candidate surrogate promoters. Accordingly,

5 depending on the desired expression level needed for a specific gene product, a particular identified surrogate promoter can be selected using this assay. For example, if simply the highest expression level is desired, then the candidate promoter that produces the largest colorimetric change may be selected. If a lower level of expression is desired, for example, because the intended product to be expressed is toxic at high levels

10 or must be expressed at equivalent levels with another product, a weaker surrogate promoter can be identified, selected, and used as described.

The plasmid pLOI2311 contains the *celY* coding region under the control of the *lac* promoter and pLOI2316 was identified as a clone oriented to express *celY* from the *lac* promoter as determined by endoglucanase indicator plates. Replacement of the

15 native promoter with the *lac* promoter increased *celY* expression by recombinant *E.coli* harboring this plasmid. In addition, to minimize problems associated with the expression of heterologous genes in industrial strains such as the ethanologenic *K. oxytoca* P2 strain, unregulated promoters were isolated from random fragments of *Z. mobilis* DNA using functional assays. Using this method, a plasmid containing a *Z.*

20 *mobilis Sau3A1* fragment as a heterologous was constructed (pLOI2323) (see example 4 and Figures 13 and 14). In addition, high levels of EGZ were produced by *E.coli* harboring the plasmid pLOI1620. See examples 3 and 4 for more details concerning the construction of the plasmids and selection of heterologous promoters.

Example 4 describes the construction of the *celY*, *celZ* integration vector with

25 surrogate promoters (pLOI2352) (See Figure 15). To ensure the stability of this plasmid, hybrid genes were integrated into the chromosome, and the antibiotic resistance markers used in construction were deleted using the FLP recombinase system (Martinez-Morales, *et al.* (1999) *J. Bacteriol.* 181:7143-7148) to facilitate further genetic modifications.

30

### ***III. Methods of Use***

#### Degrading or Depolymerizing a Complex Saccharide

In one embodiment, the host cell of the invention is used to degrade or depolymerize a complex sugar, *e.g.*, lignocellulose or an oligosaccharide into a smaller

35 sugar moiety. To accomplish this, the host cell of the invention preferably expresses one or more polysaccharases, *e.g.*, endoglucanases, such as EGY and EGZ and these polysaccharases may be liberated naturally from the producer organism. Alternatively,

the polysaccharase is liberated from the producer cell by physically disrupting the cell. Various methods for mechanically (e.g., shearing, sonication), enzymatically (e.g., lysozyme), or chemically disrupting cells, are known in the art, and any of these methods may be employed. Once the desired polypeptide is liberated from the inner cell  
5 space it may be used to degrade a complex saccharide substrate into smaller sugar moieties for subsequent bioconversion into ethanol. The liberated polysaccharase may be purified using standard biochemical techniques known in the art. Alternatively, the liberated polysaccharase need not be purified or isolated from the other cellular components and can be applied directly to the sugar substrate.

10 Accordingly, it will be appreciated by the skilled artisan that one or more polysaccharases can be selected for their activity and a composition may be formulated having an optimized activity, preferably synergistic activity, for degrading a complex sugar. The composition may take e.g., the form of an unpurified, semi-purified, or purified endoglucanase activity which is mixed with one or more endoglucanase  
15 activities in a ratio that provides optimal degrading of a complex sugar. Alternatively, each enzyme activity may be separately formulated with instructions for use, i.e., mixing or applying in a preferred order and/or ratio, in order to achieve optimal degrading of a complex sugar.

In another embodiment, a host cell is employed that coexpresses one or more  
20 polysaccharases and a secretory protein/s such that the polysaccharases are secreted into the growth medium. This eliminates the above-mentioned step of having to liberate the polysaccharases from the host cell. When employing this type of host, the host may be used directly in an aqueous solution containing an oligosaccharide.

In another embodiment, a host cell of the invention is designed to express more  
25 than one polysaccharase or is mixed with another host expressing a different polysaccharase in a ratio sufficient for the synergistic degradation of an oligosaccharide to occur. For example, one host cell could express a heterologous endoglucanase (e.g., EGY) while another host cell could express another endoglucanase (e.g., EGZ), and these cells could be combined to form a heterogeneous culture having synergistic  
30 activity in the degradation of oligosaccharides. Alternatively, in a preferred embodiment, a single host strain is engineered to produce all of the above polysaccharases. In either case, a culture of recombinant host/s is produced having high expression of the desired polysaccharases for application to an oligosaccharide. If desired, this mixture can be combined with an additional cellulase, e.g., an exogenous  
35 cellulase, such as a fungal cellulase. This mixture is then used to degrade a complex substrate. Alternatively, prior to the addition of the complex sugar substrate, the

polysaccharase/s are purified from the cells and/or media using standard biochemical techniques and used as a pure enzyme source for depolymerizing a sugar substrate.

It will be appreciated by the skilled artisan, that the ethanol-producing bacterial strains of the invention are superior hosts for the production of recombinant proteins because, under anaerobic conditions (e.g., in the absence of oxygen), there is less opportunity for improper folding of the protein (e.g., due to inappropriate disulfide bond formation). Thus, the hosts and culture conditions of the invention potentially result in the greater recovery of a biologically active product.

10 Fermenting a Complex Saccharide

In a preferred embodiment of the present invention, the host cell having the above mentioned attributes is also ethanologenic. Accordingly, such a host cell can be applied in synergistically degrading or depolymerizing a complex saccharide into a monosaccharide. Subsequently, the cell can catabolize the simpler sugar into ethanol by fermentation. This process of concurrent complex saccharide depolymerization into smaller sugar residues followed by fermentation is referred to as simultaneous saccharification and fermentation (SSF).

Typically, fermentation conditions are selected that provide an optimal pH and temperature for promoting the best growth kinetics of the producer host cell strain and catalytic conditions for the enzymes produced by the culture (Doran *et al.*, (1993) *Biotechnol. Progress.* 9:533-538). For example, for *Klebsiella*, e.g., the P2 strain, optimal conditions were determined to be between 35-37° C and pH 5.0- pH 5.4. Under these conditions, even exogenously added fungal endoglucanases and exoglucanases are quite stable and continue to function for long periods of time. Other conditions are discussed in the Examples. Moreover, it will be appreciated by the skilled artisan, that only routine experimentation is needed, using techniques known in the art, for optimizing a given fermentation reaction of the invention.

Currently, the conversion of a complex saccharide such as lignocellulose, is a very involved, multi-step process. For example, the lignocellulose must first be degraded or depolymerized using acid hydrolysis. This is then followed by steps that separate liquids from solids and these products are subsequently washed and detoxified to result in cellulose that can be further depolymerized (using added cellulases) and finally, fermented by a suitable ethanologenic host cell. In contrast, the fermenting of corn is much simpler in that amylases can be used to break down the corn starch for immediate bioconversion by an ethanologenic host in essentially a one-step process.

Accordingly, it will be appreciated by the skilled artisan that the recombinant hosts and methods of the invention afford the use of a similarly simpler and more

efficient process for fermenting lignocellulose. For example, the method of the invention is intended to encompass a method that avoids acid hydrolysis altogether. Moreover, the hosts of the invention have the following advantages, 1) efficiency of pentose and hexose co-fermentation; 2) resistance to toxins; 3) production of enzymes 5 for complex saccharide depolymerization; and 4) environmental hardiness. Therefore, the complexity of depolymerizing lignocellulose can be simplified using an improved biocatalyst of the invention. Indeed, in one preferred embodiment of the invention, the reaction can be conducted in a single reaction vessel and in the absence of acid hydrolysis, *e.g.*, as an SSF process.

10

#### Potential Substrates for Bioconversion into Ethanol

One advantage of the invention is the ability to use a saccharide source that has been, heretofore, underutilized. Consequently, a number of complex saccharide substrates may be used as a starting source for depolymerization and subsequent 15 fermentation using the host cells and methods of the invention. Ideally, a recyclable resource may be used in the SSF process. Mixed waste office paper is a preferred substrate (Brooks *et al.*, (1995) *Biotechnol. Progress.* 11:619-625; Ingram *et al.*, (1995) U.S.P.N. 5,424,202), and is much more readily digested than acid pretreated bagasse (Doran *et al.*, (1994) *Biotech. Bioeng.* 44:240-247) or highly purified crystalline 20 cellulose (Doran *et al.* (1993) *Biotechnol. Progress.* 9:533-538). Glucanases, both endoglucanases and exoglucanases, contain a cellulose binding domain, and these enzymes can be readily recycled for subsequent fermentations by harvesting the undigested cellulose residue using centrifugation (Brooks *et al.*, (1995) *Biotechnol.* 25 *Progress.* 11:619-625). By adding this residue with bound enzyme as a starter, ethanol yields (per unit substrate) were increased to over 80% of the theoretical yield with a concurrent 60% reduction in fungal enzyme usage (Figure 2). Such approaches work well with purified cellulose, although the number of recycling steps may be limited with substrates with a higher lignin content. Other substrate sources that are within the scope 30 of the invention include any type of processed or unprocessed plant material, *e.g.*, lawn clippings, husks, cobs, stems, leaves, fibers, pulp, hemp, sawdust, newspapers, *etc.*

This invention is further illustrated by the following examples, which should not be construed as limiting.

**EXAMPLE 1****Methods for Making Recombinant *Escherichia* Hosts Suitable for Fermenting Oligosaccharides into Ethanol**

- In this example, methods for developing and using *Escherichia* hosts suitable for fermenting oligosaccharides into ethanol are described. In particular, a strong promoter is identified which can be used to increase the expression of a polysaccharase (e.g., glucanase). In addition, genes from *Erwinia chrysanthemi* are employed to facilitate polysaccharase secretion thereby eliminating the need for cell disruption in order to release the desired polysaccharase activity.
- Throughout this example, the following materials and methods are used unless otherwise stated.

***Materials and Methods******Organisms and Culture Conditions***

- The bacterial strains and plasmids used in this example are listed in Table 1, below.

For plasmid constructions, the host cell *E. coli* DH5 $\alpha$  was used. The particular gene employed encoding a polysaccharase (e.g., glucanase) was the *celZ* gene derived from *Erwinia chrysanthemi* P86021 (Beall, (1995) Ph.D. Dissertation, University of Florida; Wood *et al.*, (1997) *Biotech. Bioeng.* 55:547-555). The particular genes used for improving secretion were the *out* genes derived from *E. chrysanthemi* EC16 (He *et al.*, (1991) *Proc. Natl. Acad. Sci. USA*. 88:1079-1083).

Typically, host cell cultures were grown in Luria-Bertani broth (LB) ( $10\text{ g L}^{-1}$  Difco<sup>®</sup> tryptone,  $5\text{ g L}^{-1}$  Difco<sup>®</sup> yeast extract,  $5\text{ g L}^{-1}$  sodium chloride) or on Luria agar (LB supplemented with  $15\text{ g L}^{-1}$  of agar). For screening host cells having glucanase *celZ* activity (EGZ), CMC-plates (Luria agar plates containing carboxymethyl cellulose ( $3\text{ g L}^{-1}$ )) were used (Wood *et al.*, (1988) *Methods in Enzymology* 160:87-112). When appropriate, the antibiotics ampicillin ( $50\text{ mg L}^{-1}$ ), spectinomycin ( $100\text{ g L}^{-1}$ ), kanamycin ( $50\text{ g L}^{-1}$ ) were added to the media for selection of recombinant or integrant host cells containing resistance markers. Constructs containing plasmids with a temperature conditional pSC101 replicon (Posfai *et al.*, (1997) *J. Bacteriol.* 179:4426-4428) were grown at  $30^\circ\text{C}$  and, unless stated otherwise, constructs with pUC-based plasmids were grown at  $37^\circ\text{C}$ .

**TABLE 1. Strains and Plasmids Used**

Strains/Plasmids	Description	Sources/References
<b>Strains</b>		
<i>Z. mobilis</i> CP4	Prototrophic	Osman, et al., (1985) <i>J. Bact.</i> 164:173-180
<i>E. coli</i> strain DH5 $\alpha$	<i>lacZ M15 recA</i>	Bethesda Research Laboratory
<i>E. coli</i> strain B	Prototrophic	ATCC 11303
<i>E. coli</i> strain HB 101	<i>RecA lacY recA</i>	ATCC 37159
<b>Plasmids</b>		
pUC19	<i>bla</i> cloning vector	New England Biolabs
pST76-K	<i>kan</i> low copy number, temp. sensitive	Posfai, et al., (1997) <i>J. Bacteriol.</i> 179:4426-4428
pRK2013	<i>kan</i> mobilizing helper plasmid ( <i>mob</i> +)	ATCC
pCPP2006	Sp <sup>r</sup> , ca. 40 kbp plasmid carrying the complete <i>out</i> genes from <i>E. chrysanthemi</i> EC16	He, et al., (1991) <i>P.N.A.S.</i> 88:1079-1083
pLOI1620	<i>bla celZ</i>	Beall, et al., (1995) Ph.D. Dissertation, U. of Florida
pLOI2164	pLOI1620 with <i>Bam</i> HI site removed (Klenow)	See text
pLOI2170	<i>Nde</i> I- <i>Hind</i> III fragment (promoterless <i>celZ</i> ) from pLOI2164 cloned into pUC19	See text
pLOI2171	<i>Bam</i> HI- <i>Sph</i> I fragment (promoterless <i>celZ</i> ) from pLOI2170 cloned into pST76-K	See text
pLOI2173	<i>Eco</i> RI- <i>Sph</i> I fragment ( <i>celZ</i> with native promoter) from pLOI2164 cloned into pST76-K	See text
pLOI2174	<i>Eco</i> RI- <i>Bam</i> HI fragment ( <i>gap</i> promoter) cloned into pLOI2171	See text
pLOI2175	<i>Eco</i> RI- <i>Bam</i> HI fragment ( <i>eno</i> promoter) cloned into pLOI2171	See text
pLOI2177	Random <i>Sau</i> 3A1 <i>Z. mobilis</i> DNA fragment cloned into pLOI2171	See text
pLOI2178	Random <i>Sau</i> 3A1 <i>Z. mobilis</i> DNA fragment cloned into pLOI2171	See text
pLOI2179	Random <i>Sau</i> 3A1 <i>Z. mobilis</i> DNA fragment cloned into pLOI2171	See text
pLOI2180	Random <i>Sau</i> 3A1 <i>Z. mobilis</i> DNA fragment cloned into pLOI2171	See text
pLOI2181	Random <i>Sau</i> 3A1 <i>Z. mobilis</i> DNA fragment cloned into pLOI2171	See text
pLOI2182	Random <i>Sau</i> 3A1 <i>Z. mobilis</i> DNA fragment cloned into pLOI2171	See text
pLOI2183	Random <i>Sau</i> 3A1 <i>Z. mobilis</i> DNA fragment cloned into pLOI2171	See text
pLOI2184	Random <i>Sau</i> 3A1 <i>Z. mobilis</i> DNA fragment cloned into pLOI2171	See text
pLOI2196	pLOI2177 fused into pUC19 at the <i>Pst</i> I site	See text
pLOI2197	pLOI2180 fused into pUC19 at the <i>Pst</i> I site	See text
pLOI2198	pLOI2182 fused into pUC19 at the <i>Pst</i> I site	See text
pLOI2199	pLOI2183 fused into pUC19 at the <i>Pst</i> I site	See text
pLOI2307	<i>Eco</i> RI- <i>Sph</i> I fragment from pLOI2183 cloned into pUC19	See text

### *Genetic Methods*

Standard techniques were used for all plasmid constructions (Ausubel *et al.*, (1987) *Current Protocols in Molecular Biology*, John Wiley & Sons, Inc.; Sambrook *et al.*, (1989) *Molecular cloning: a laboratory manual*, 2<sup>nd</sup> ed. C.S.H.L., Cold Spring Harbor, N.Y). For conducting small-scale plasmid isolation, the TELT procedure was performed. For large-scale plasmid isolation, the Promega® Wizard Kit was used. For isolating DNA fragments from gels, the Qiaquick® Gel Extraction Kit from Qiagen® was employed. To isolate chromosomal DNA from *E. coli* and *Z. mobilis* the methods of Cutting and Yomano were used (Cutting *et al.*, (1990), Genetic analysis, pp. 61-74, In, Molecular biological methods for *Bacillus*, John Wiley & Sons, Inc.; Yomano *et al.*, (1993) *J. Bacteriol.* 175:3926-3933).

To isolate the two glycolytic gene promoters (*e.g.*, *gap* and *eno*) described herein, purified chromosomal DNA from *E. coli* DH5 $\alpha$  was used as a template for the PCR (polymerase chain reaction) amplification of these nucleic acids using the following primer pairs: *gap* promoter, 5'-CGAATTCTGCCGAAGTTATTAGCCA-3' (SEQ ID NO: 3) and 5'-AAGGATCCTTCCACCACTAGCTATTGTTAGTGA-3' (SEQ ID NO: 4); *eno* promoter, 5'-AGAATTCTGCCAGTTGGTGACGATAG-3' (SEQ ID NO: 5) and 5'-CAGGATCCCCTCAAGTCACTAGTTAACTG-3' (SEQ ID NO: 6). The *out* genes encoding secretory proteins derived from *E. chrysanthemi* (pCPP2006) were conjugated into *E. coli* using pRK2013 for mobilization (Figurski *et al.*, (1979) *Proc. Natl. Acad. Sci. USA*. 76: 1648-1652; Murata *et al.*, (1990) *J. Bacteriol.* 172:2970-2978).

To determine the sequence of various DNAs of interest, the dideoxy sequencing method using fluorescent primers was performed on a LI-COR Model 4000-L DNA Sequencer. The pST76-K-based plasmids were sequenced in one direction using a T7 primer (5'-TAATACGACTCACTATAGGG-3' (SEQ ID NO: 7)). The pUC18- and pUC19-based plasmids were sequenced in two directions using either a forward primer (5'-CACGACGTTGTAAAACGAC-3' (SEQ ID NO: 8)) or a reverse primer (5'-TAACAATTTCACACAGGA-3' (SEQ ID NO: 9)). The extension reactions of the sequencing method were performed using a Perkin Elmer GeneAmp® PCR 9600 and SequiTherm Long-Read Sequencing Kit-LC®. Resultant sequences were subsequently analyzed using the Wisconsin Genetic Computer Group (GCG) software package (Devereux *et al.*, (1984) *Nucleic Acids Rev.* 12:387-395).

To determine the start of transcriptional initiation in the above-mentioned promoters, primer extension analysis was performed using standard techniques. In particular, promoter regions were identified by mapping the transcriptional start sites using a primer finding correspondence within the *celZ* gene RNA that was isolated from

cells in late exponential phase using a Qiagen RNeasy® kit. Briefly, cells were treated with lysozyme (400 µg/ml) in TE (Tris-HCl, EDTA) containing 0.2 M sucrose and incubated at 25° C for 5 min prior to lysis. Liberated RNA was subjected to ethanol precipitation and subsequently dissolved in 20 µl of Promega® AMV reverse transcriptase buffer (50 mM Tris-HCl, pH 8.3, 50 mM KCl, 10 mM MgCl<sub>2</sub>, 0.5 mM spermidine, 10 mM DTT). An IRD41-labeled primer (5'-GACTGGATGGTTATCCGAATAAGAGAGAGG-3' (SEQ ID NO: 10)) from LI-Cor Inc. was then added and the sample was denatured at 80° C for 5 min, annealed at 55° C for 1 hr, and purified by alcohol precipitation. Annealed samples were dissolved in 19 µl of AMV reverse transcriptase buffer containing 500 µM dNTPs and 10 units AMV reverse transcriptase, and incubated for extension (1 h at 42°C). Products were treated with 0.5 µg/ml DNase-free RNase A, precipitated, dissolved in loading buffer, and compared to parallel dideoxy promoter sequences obtained using the LI-COR Model 4000-L DNA sequencer.

15

#### *Polysaccharase Activity*

To determine the amount of polysaccharase activity (e.g., glucanase activity) resulting from expression of the *celZ* gene, a Congo Red procedure was used (Wood *et al.*, (1988) *Methods in Enzymology* 160:87-112). In particular, selected clones were transferred to gridded CMC plates and incubated for 18 h at 30° C and then stained and recombinant host cells expressing glucanase formed yellow zones on a red background. Accordingly, the diameters of these colorimetric zones were recorded as a relative measure of *celZ* expression.

Glucanase activity (EGZ) was also measured using carboxymethyl cellulose as a substrate. In this test, appropriate dilutions of cell-free culture broth (extracellular activity) or broth containing cells treated with ultrasound (total activity) were assayed at 35° C in 50 mM citrate buffer (pH 5.2) containing carboxymethyl cellulose (20 g L<sup>-1</sup>). Conditions for optimal enzyme release for 3-4 ml samples were determined to be 4 pulses at full power for 1 second each using a cell disruptor (Model W-220F, Heat System-Ultrasonics Inc., Plainview, NY). To stop the enzyme reactions of the assay, samples were heated in a boiling water bath for 10 min. To measure reducing sugars liberated enzymatically by the glucanase, a dinitrosalicylic acid reagent was employed using glucose as a standard (Wood *et al.*, (1988) *Methods in Enzymology* 160:87-112). The amount of enzyme activity (IU) was expressed as µmols of reducing sugar released per min or as a percentage of total activity from an average of two or more determinations.

TOP SECRET//SI//REL TO USA, UK, FVEY

### *Ultrastructural Analysis*

To determine the ultrastructure of various recombinant host cells, fresh colonies from Luria agar plates were prepared for analysis by fixing in 2% glutaraldehyde in 0.2 M sodium cacodylate buffer (pH 7) followed by incubation in 1% osmium tetroxide and 5 followed by 1% uranyl acetate in distilled water. Samples were dehydrated in ethanol, embedded in Spurr's plastic, and ultrathin sections were prepared and examined using a Zeiss® EM-IOCA electron microscope (Spur (1969) *J. Ultrastruct. Res.* 26:31).

### *Construction of a Low Copy Promoter Probe Vector Using celZ as the Reporter Gene*

10 To facilitate the isolation of strong promoters, a low copy vector was constructed with a pSC101 replicon and a *Bam*HI site immediately preceding a promoterless *celZ* gene (pLOI2171). Accordingly, this promoterless plasmid was used as a negative control. The plasmid pLOI1620 was used as a source of *celZ* and is a pUC18 derivative with expression from consecutive *lac* and *celZ* promoters. The *Bam*HI site in this  
15 plasmid was eliminated by digestion and Klenow treatment (pLOI2164). The *celZ* gene was isolated as a promoterless *Nde*I fragment after Klenow treatment. The resulting blunt fragment was digested with *Hind*III to remove downstream DNA and ligated into pUC19 (*Hind*III to *Hinc*II) to produce pLOI2170. In this plasmid, *celZ* is oriented opposite to the direction of *lacZ* transcription and was only weakly expressed. The  
20 *Bam*HI (amino terminus)-*Sph*I (carboxyl terminus) fragment from pLOI2170 containing *celZ* was then cloned into the corresponding sites of pST76-K, a low copy vector with a temperature sensitive replicon, to produce pLOI2171 (Fig. 3). Expression of *celZ* in this vector was extremely low facilitating its use as a probe for candidate strong promoters.

### *Analysis of celZ Expression from Two E. coli Glycolytic Promoters (gap and eno)*

25 Two exemplary promoters driving glycolytic genes (*gap* and *eno*) in *E. coli* were examined for their ability to drive the expression of the heterologous *celZ* gene encoding glucanase. Chromosomal DNA from the *E. coli* DH5 $\alpha$  strain was used as a template to amplify the *gap* and *eno* promoter regions by the polymerase chain reaction. The  
30 resulting fragments of approximately 400 bp each were digested with *Eco*RI and *Bam*HI and cloned into the corresponding sites in front of a promoterless *celZ* gene in pLOI2171 to produce pLOI2174 (*gap* promoter) and pLOI2175 (*eno* promoter). As a control, the *Eco*RI-*Sph*I fragment from pLOI2164 containing the complete *celZ* gene and native *E. chrysanthemi* promoter was cloned into the corresponding sites of  
35 pST76-K to produce pLOI2173. These three plasmids were transformed into *E. coli* strains B and DH5 $\alpha$  and glucanase activity (EGZ) was compared. For both strains of *E. coli*, glucanase activities were lower on CMC plates with *E. coli* glycolytic promoters

- than with pLOI2173 containing the native *E. chrysanthemi* promoter (Table 2). Assuming activity is related to the square of the radius of each zone (Fick's Law of diffusion), EGZ production with glycolytic promoters (pLOI2174 and pLOI2175) was estimated to be 33% to 65% lower than in the native promoter construct (pLOI2373).
- 5 Accordingly, other candidate promoters for driving high levels of *celZ* gene expression were investigated.

***Identifying and Cloning Random DNA Fragments Suitable for Use as Promoters for Heterologous Gene Expression***

- 10 Random fragments derived from *Z. mobilis* can be an effective source of surrogate promoters for the high level expression of heterologous genes in *E. coli*. (Conway *et al.*, (1987) *J. Bacteriol.* 169:2327-2335; Ingram *et al.*, (1988) *Appl. Environ. Micro.* 54:397-404). Accordingly, to identify surrogate promoters for *Erwinia celZ* expression, *Z. mobilis* chromosomal DNA was extensively digested with *Sau3AI* and
- 15 resulting fragments were ligated into pLOI2171 at the *BamHI* site and transformed into *E. coli* DH5 $\alpha$  to generate a library of potential candidate promoters. To rapidly identify superior candidate promoters capable of driving *celZ* gene expression in *E. coli*, the following biological screen was employed. Colonies transformed with *celZ* plasmids having different random candidate promoters were transferred to gridded CMC plates
- 20 and stained for glucanase activity after incubation (Table 2). Approximately 20% of the 18,000 clones tested were CMC positive. The 75 clones which produced larger zones than the control, pLOI2173, were examined further using another strain, *E. coli* B.

**TABLE 2. Evaluation of promoter strength for *celZ* expression in *E. coli* using CMC indicator plates.**

Plasmids	<i>E. coli</i> DH5 $\alpha$ host			<i>E. coli</i> B host		
	Number of Plasmids <sup>a</sup>	CMC zone diameter (mm) <sup>b</sup>	% of native promoter ( $100 \cdot R_x^2/R_c^2$ ) <sup>c</sup>	Number of plasmids	CMC zone diameter (mm)	% of native promoter ( $100 \cdot R_x^2/R_c^2$ )
pLOI2171 (promoterless)	1	0	--	--	--	--
pLOI2173 (native promoter)	1	5.0	100	1	4.5	100
pLOI2174 (gap promoter)	1	4.0	77	1	3.5	60
pLOI2175 ( <i>eno</i> promoter)	1	3.0	43	1	2.8	35
<i>Z. mobilis</i> promoters						
Group I	5	13.0	676	4	10.8-11.3	570-625
Group II	14	9.0-11.0	324-484	17	9.0-10.5	445-545
Group III	56	6.0-9.0	144-324	54	5.0-8.8	125-375

<sup>a</sup> The number of clones which the indicated range of activities.

<sup>b</sup> The average size of the diameters from three CMC digestion zones.

<sup>c</sup>  $R_x^2$  is the square of the radius of the clear zone with the test plasmid;  $R_c^2$  is the square of the radius of the clear zone for the control (pLOI2173).

5

10

Thus, promoter strength for selected candidate promoters was confirmed in two different strains with, in general, recombinants of DH5 $\alpha$  producing larger zones (e.g., more glucanase) than recombinants of strain B. However, relative promoter strength in each host was similar for most clones. Based on these analyses of glucanase production as measured by zone size using CMC plates, four clones appeared to express *celZ* at approximately 6-fold higher levels than the construct with the original *E. chrysanthemi* *celZ* gene (pLOI2173), and at 10-fold higher levels than either of the *E. coli* glycolytic promoters. Accordingly, these and similarly strong candidate promoters were selected for further study.

15

#### **Production and Secretion of Glucanase**

Eight plasmid derivatives of pST76-K (pLOI2177 to pLOI2184) were selected from the above-described screen (see Group I and Group II (Table 2)) and assayed for total glucanase activity in *E. coli* strain B (Table 3). The four plasmids giving rise to the largest zones on CMC plates were also confirmed to have the highest glucanase activities (pLOI2177, pLOI2180, pLOI2182, and pLOI2183). The activities were approximately 6-fold higher than that of the unmodified *celZ* (pLOI2173), in excellent agreement with our estimate using the square of the radius of the cleared zone on CMC plates. Figure 4 shows a comparison of activity estimates from CMC plates and *in vitro*

20

25

enzyme assays for strain B containing a variety of different promoters, with and without the addition of *out* genes encoding secretory proteins. Although there is some scatter, a direct relationship is clearly evident which validates the plate method for estimating relative activity. The original construct in pUC18, a high copy plasmid, was also included for comparison (pLOI2164). This construct with consecutive *lac* and *celZ* promoters produced less EGZ activity than three of the low copy plasmids with surrogate promoters (pLOI2177, pLOI2182, and pLOI2183). Thus, to increase *celZ* expression of glucanase even more, the DNA fragment containing *celZ* and the most effective surrogate promoter was isolated from pLOI2183 (as a *Eco*RI-*Sph*I fragment) and inserted into pUC19 with transcription oriented opposite to that of the *lac* promoter (pLOI2307). Accordingly, the above-identified strong surrogate promoter when incorporated into a high copy plasmid, further increased glucanase activity by 2-fold.

### ***Engineering Increased Secretion of Glucanase***

To further improve on the above-described results for increasing expression of *celZ* encoded glucanase, the above host cells were engineered for increased secretion. Genes encoding secretory proteins (*e.g.*, the *out* genes) derived from *E. chrysanthemi* EC16 were used for improving the export of the glucanase using the plasmid as described in He *et al.* that contains *out* genes (pCPP2006) (He *et al.*, (1991) *Proc. Natl. Acad. Sci. USA.* 88:1079-1083). The increased secretion of EGZ in *E. coli* B was investigated and results are presented in Table 3.

**TABLE 3. Comparison of promoters for EGZ production and secretion in *E. coli* B**

Plasmids <sup>a</sup>	Without secretion genes		With secretion genes (pCPP2006)	
	Total activity (IU/L) <sup>b</sup>	Extracellular <sup>c</sup> (%)	Total Activity (IU/L)	Extracellular <sup>c</sup> (%)
pLOI2173	620	17	1,100	43
pLOI2177	3,700	10	5,500	44
pLOI2178	2,200	9	3,500	49
pLOI2179	2,000	10	3,000	50
pLOI2180	2,900	8	6,300	39
pLOI2181	1,800	11	4,100	46
pLOI2182	3,500	7	6,600	38
pLOI2183	3,400	7	6,900	39
pLOI2184	2,100	12	2,400	39
pLOI2164	3,200	20	6,900	74
pLOI2307	6,600	28	13,000	60

<sup>a</sup> Plasmids pLOI2173 and pLOI2164 contain the *celZ* native promoter; pLOI2307 contains the strong promoter from pLOI2183.

<sup>b</sup> Plasmids pLOI2164 and pLOI2307 are pUC-based plasmids (high copy number). All other plasmids are derivatives of pST76-K (low copy number).

<sup>c</sup> Glucanase activities were determined after 16 h of growth at 30°C.

<sup>c</sup> Extracellular activity (secreted or released).

Recombinant hosts with low copy plasmids produced only 7- 17% of the total EGZ extracellularly (after 16 hours of growth) without the additional heterologous secretory proteins (*out* proteins encoded by plasmid pCPP2006). A larger fraction of 5 EGZ (20-28%) was found in the extracellular broth surrounding host cells with the high-copy pUC-based plasmids than with the low copy pST76-based plasmids containing the same promoters. However, in either case, the addition of *out* genes encoding secretory proteins (e.g., pCPP2006) increased the total level of expression by up to 2-fold and increased the fraction of extracellular enzyme (38-74%) by 10 approximately 4-fold. The highest activity, 13,000 IU/L of total glucanase of which 7,800 IU/L was found in the cell-free supernatant was produced by strain B having both pLOI2307 encoding *celZ* driven by a strong surrogate promoter and pCPP2006 encoding *out* secretory proteins).

It has been reported that under certain conditions (pH 7, 37° C), the specific activity for pure EGZ enzyme is 419 IU (Py *et al.*, (1991) *Protein Engineering* 4:325-333) and it has been determined that EGZ produced under these conditions is 25% more active than under the above-mentioned conditions (pH 5.2 citrate buffer, 35° C). Accordingly, assuming a specific activity of 316 IU for pure enzyme at pH 5.2 (35°C), the cultures of *E. coli* B (containing pLOI2307 and pCPP2006, *e.g.*, plasmids encoding glucanase and secretory proteins), produced approximately 41 mg of active EGZ per liter or 4-6% of the total host cell protein was active glucanase.

## *Sequence Analysis of the Strongest Promoter Derived from Z. mobilis*

The sequences of the four strongest surrogate promoters (pLOI2177, pLOI2180, pLOI2182, and pLOI2183) were determined. To facilitate this process, each was fused with pUC19 at the *Pst*I site. The resulting plasmids, pLOI2196, pLOI2197, pLOI2198, and pLOI2199, were produced at high copy numbers (ColEI replicon) and could be sequenced in both directions using M13 and T7 sequencing primers. All four plasmids contained identical pieces of *Z. mobilis* DNA and were siblings. Each was 1417 bp in length and contained 4 internal Sau3AI sites. DNA and translated protein sequences (six reading frames) of each piece were compared to the current data base. Only one fragment (281 bp internal fragment) exhibited a strong match in a BLAST search (National Center for Biotechnology Information; <http://www.ncbi.nlm.nih.gov/BLAST/>) and this fragment was 99% identical in DNA sequence to part of the *Z. mobilis* *hpnb* gene which is proposed to function in cell envelope biosynthesis (Reipen *et al.*, (1995) *Microbiology* 141:155-161). Primer extension analysis revealed a single major start site, 67 bp upstream from the *Sau*3AI/*Bam*HI junction site with *celZ*, and a second minor

start site further upstream (Fig. 5). Sequences in the -10 and -35 regions were compared to the conserved sequences for *E. coli* sigma factors (Wang *et al.*, (1989) *J. Bacteriol.* 180:5626-5631; Wise *et al.*, (1996) *J. Bacteriol.* 178:2785-2793). The dominant promoter region (approximately 85% of total start site) appears similar to a sigma<sup>70</sup> 5 promoter while the secondary promoter site resembles a sigma<sup>38</sup> promoter.

#### ***Microscopic Analysis of Recombinant Host Cells Producing Glucanase***

Little difference in cell morphology was observed between recombinants and the parental organism by light microscopy. Under the electron microscope, however, small 10 polar inclusion bodies were clearly evident in the periplasm of strain B (pLOI2164) expressing high amounts of glucanase and these inclusion bodies were presumed to contain EGZ (Fig. 6). In the strain B (pLOI2307) that produced 2-fold higher glucanase activity the inclusion bodies were even larger and occupied up to 20% of the total cell volume. The large size of these polar bodies suggests that glucanase activity 15 measurements may underestimate the total EGZ production. Typically, polar inclusion bodies were smaller in host cells also having constructs encoding the *out* secretory proteins, which allow for increased secretion of proteins from the periplasmic space. As expected, no periplasmic inclusion bodies were evident in the negative control strain B (pUC19) which does not produce glucanase.

20

RECORDED BY SENSER

## EXAMPLE 2

### Recombinant *Klebsiella* Hosts Suitable for Fermenting Oligosaccharides into Ethanol

In this example, a recombinant *Klebsiella* host, suitable for use as a biocatalyst for depolymerizing and fermenting oligosaccharides into ethanol, is described.

Throughout this example, the following materials and methods are used unless otherwise stated.

#### **Materials and Methods**

##### *Bacteria, Plasmids, and Culture Conditions*

The strains and plasmids that were used in this exemplification are summarized in Table 4 below.

**TABLE 4. Strains and Plasmids Used**

Strains/Plasmids	Properties	Sources/References
<b>Strains</b>		
<i>Zymomonas mobilis</i> CP4	Prototrophic	Ingram <i>et al.</i> (1988) <i>Appl. Environ. Micro.</i> 54:397-404
<i>Escherichia coli</i>		
DH5 $\alpha$	<i>lacZ</i> M15 <i>recA</i>	Bethesda Research Laboratory
HB101	<i>recA</i> <i>lacY</i> <i>recA</i>	ATCC 37159
<i>Klebsiella oxytoca</i>		
M5A1	Prototrophic	Wood <i>et al.</i> (1992) <i>Appl. Environ. Micro.</i> 58:2103-2110
P2	<i>pfl</i> :: <i>pdc adhB cat</i>	Wood <i>et al.</i> (1992) <i>Appl. Environ. Micro.</i> 58:2103-2110
SZ1	<i>pfl</i> :: <i>pdc adhB cat</i> ; integrated <i>celZ</i> ; <i>tet</i>	See text
SZ2	<i>pfl</i> :: <i>pdc adhB cat</i> ; integrated <i>celZ</i> ; <i>tet</i>	See text
SZ3	<i>pfl</i> :: <i>pdc adhB cat</i> ; integrated <i>celZ</i> ; <i>tet</i>	See text
SZ4	<i>pfl</i> :: <i>pdc adhB cat</i> ; integrated <i>celZ</i> ; <i>tet</i>	See text
SZ5	<i>pfl</i> :: <i>pdc adhB cat</i> ; integrated <i>celZ</i> ; <i>tet</i>	See text
SZ6	<i>pfl</i> :: <i>pdc adhB cat</i> ; integrated <i>celZ</i> ; <i>tet</i>	See text
SZ7	<i>pfl</i> :: <i>pdc adhB cat</i> ; integrated <i>celZ</i> ; <i>tet</i>	See text
SZ8	<i>pfl</i> :: <i>pdc adhB cat</i> ; integrated <i>celZ</i> ; <i>tet</i>	See text
SZ9	<i>pfl</i> :: <i>pdc adhB cat</i> ; integrated <i>celZ</i> ; <i>tet</i>	See text
SZ10	<i>pfl</i> :: <i>pdc adhB cat</i> ; integrated <i>celZ</i> ; <i>tet</i>	See text
<b>Plasmids</b>		
pUC19	<i>bla</i> cloning vector	New England Biolabs
pBR322	<i>bla</i> <i>tet</i> cloning vector	New England Biolabs
pLOI1620	<i>bla celZ</i>	Wood <i>et al.</i> (1997) <i>Biotech. Bioeng.</i> 55:547-555
pRK2013	<i>kan</i> mobilizing helper plasmid ( <i>mob</i> <sup>+</sup> )	ATCC
pCPP2006	Sp <sup>r</sup> , 40 kbp fragment containing <i>out</i> genes from <i>E. chrysanthemi</i> EC16	He <i>et al.</i> (1991) <i>P.N.A.S.</i> 88:1079-1083
pST76-K	<i>kan</i> low copy vector containing temperature sensitive pSC101 replicon	Posfai <i>et al.</i> (1997) <i>J. Bact.</i> 179:4426-4428
pLOI2164	<i>bla celZ</i> ( <i>Bam</i> HI eliminated from pLOI1620)	See text
pLOI2173	<i>kan celZ</i> (native <i>celZ</i> promoter)	See text

**TABLE 4. Strains and Plasmids Used (Continued)**

<b>Strains/Plasmids</b>	<b>Properties</b>	<b>Sources/References</b>
<b>Plasmids (Continued)</b>		
pLOI2177	<i>kan celZ</i> (surrogate promoter from <i>Z. mobilis</i> )	See text
pLOI2178	<i>kan celZ</i> (surrogate promoter from <i>Z. mobilis</i> )	See text
pLOI2179	<i>kan celZ</i> (surrogate promoter from <i>Z. mobilis</i> )	See text
pLOI2180	<i>kan celZ</i> (surrogate promoter from <i>Z. mobilis</i> )	See text
pLOI2181	<i>kan celZ</i> (surrogate promoter from <i>Z. mobilis</i> )	See text
pLOI2182	<i>kan celZ</i> (surrogate promoter from <i>Z. mobilis</i> )	See text
pLOI2183	<i>kan celZ</i> (surrogate promoter from <i>Z. mobilis</i> )	See text
pLOI2184	<i>kan celZ</i> (surrogate promoter from <i>Z. mobilis</i> )	See text
<b>Strains/Plasmids</b>	<b>Properties</b>	<b>Sources/References</b>
pLOI2185	<i>kan celZ</i> (surrogate promoter from <i>Z. mobilis</i> )	See text
pLOI2186	<i>kan celZ</i> (surrogate promoter from <i>Z. mobilis</i> )	See text
pLOI2187	<i>kan celZ</i> (surrogate promoter from <i>Z. mobilis</i> )	See text
pLOI2188	<i>kan celZ</i> (surrogate promoter from <i>Z. mobilis</i> )	See text
pLOI2189	<i>kan celZ</i> (surrogate promoter from <i>Z. mobilis</i> )	See text
pLOI2190	<i>kan celZ</i> (surrogate promoter from <i>Z. mobilis</i> )	See text
pLOI2191	<i>kan celZ</i> (surrogate promoter from <i>Z. mobilis</i> )	See text
pLOI2192	<i>kan celZ</i> (surrogate promoter from <i>Z. mobilis</i> )	See text
pLOI2193	<i>kan celZ</i> (surrogate promoter from <i>Z. mobilis</i> )	See text
pLOI2194	<i>kan celZ</i> (surrogate promoter from <i>Z. mobilis</i> )	See text
pLOI2301	<i>AscI</i> linker inserted into <i>NdeI</i> site of pUC19	See text
pLOI2302	<i>AscI</i> linker inserted into <i>SapI</i> site of pLOI2301	See text
pLOI2303	<i>Aval-EcoRI</i> fragment from pBR322 inserted into <i>PstI</i> site of pLOI2302 after Klenow treatment	See text
pLOI2305	<i>EcoRI</i> DNA fragment of <i>K. oxytoca</i> M5A1 genomic DNA (ca. 2.5 kb) cloned into the <i>Smal</i> site of pLOI2303	See text
pLOI2306	<i>EcoRI-SphI</i> fragment from pLOI2183 cloned into <i>EcoRI</i> site of pLOI2305	See text

The culture conditions used for cultivating *E. coli* and *K. oxytoca* M5A1 typically employed Luria-Bertani broth (LB) containing per liter: 10 g Difco® tryptone, 5 g yeast extract, and 5 g sodium chloride, or, alternatively, Luria agar (LB supplemented with 15 g of agar) (Sambrook *et al.*, (1989), *Molecular Cloning: A*

- ## 5 *Laboratory Manual*, C.S.H.L., Cold Spring Harbor, N.Y.).

For screening bacterial colonies under selective conditions, CMC-plates (Luria agar plates containing 3 g L<sup>-1</sup> carboxymethyl cellulose) were used to determine levels of glucanase activity expressed by a given bacterial strain (Wood *et al.* (1988) *Enzymology*, 160:87-112). For cultivating ethanologenic strains, glucose was added to solid media (20 g L<sup>-1</sup>) and broth (50 g L<sup>-1</sup>). In determining glucanase activity, the glucose in the growth media was replaced with sorbitol (50 g L<sup>-1</sup>), a non-reducing sugar. For cultivating various strains or cultures in preparation for introducing nucleic acids by electroporation, a modified SOC medium was used (*e.g.*, 20 g L<sup>-1</sup> Difco<sup>®</sup> tryptone, 5 g L<sup>-1</sup>, Difco<sup>®</sup> yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM MgSO<sub>4</sub>, 10 mM MgCl<sub>2</sub>, and 50 g L<sup>-1</sup> glucose). The antibiotics ampicillin (50 mg L<sup>-1</sup>), spectinomycin (100 mg L<sup>-1</sup>), kanamycin (50 mg L<sup>-1</sup>), tetracycline (6 or 12 mg L<sup>-1</sup>), and chloramphenicol (40, 200, or 600 mg L<sup>-1</sup>) were added when appropriate for selection of recombinant hosts bearing antibiotic resistance markers. Unless stated otherwise, cultures were grown at 37° C. Ethanologenic strains and strains containing plasmids with a temperature-sensitive pSC101 replicon were grown at 30° C.

## *Genetic Methods*

For plasmid construction, cloning, and transformations, standard methods and *E. coli* DH5 $\alpha$  hosts were used (Ausubel *et al.* (1987) Current Protocols in Molecular Biology. John Wiley & Sons, Inc.; Sambrook *et al.*, (1989) *Molecular Cloning: A Laboratory Manual*, C.S.H.L., Cold Spring Harbor, N.Y.). Construction of the celZ integration vector, pLOI2306, was performed as shown in Figure 7. A circular DNA fragment lacking a replicon from pLOI2306 (see Figure 7) was electroporated into the ethanologenic *K. oxytoca* P2 using a Bio-Rad Gene Pulser using the following conditions: 2.5 kV and 25  $\mu$ F with a measured time constant of 3.8-4.0 msec (Comaduran *et al.* (1998) *Biotechnol. Lett.* 20:489-493). The *E. chrysanthemi* EC 16 secretion system (pCPP2006) was conjugated into *K. oxytoca* using pRK2013 for mobilization (Murata *et al.* (1990) *J. Bacteriol.* 172:2970-2978). Small scale and large scale plasmid isolations were performed using the TELT procedure and a Promega Wizard Kit, respectively. DNA fragments were isolated from gels using a Qiaquick® Gel Extraction Kit from Qiagen® (Qiagen Inc., Chatsworth, CA). Chromosomal DNA from *K. oxytoca* M5A1 and *Z. mobilis* CP4 were isolated as described by Cutting and

Yomano (see Example 1). The DNAs of interest were sequenced using a LI-COR Model 4000-L DNA sequencer (Wood *et al.* (1997) *Biotech. Bioeng.* 55:547-555).

#### *Chromosomal Integration of celZ*

- 5 Two approaches were employed for chromosomal integration of *celZ*, using selection with a temperature-conditional plasmid (pLOI2183) using a procedure previously described for *E. coli* (Hamilton *et al.*, (1989) *J. Bacteriol.* 171:4617-4622) and direct integration of circular DNA fragments lacking a functional replicon. This same method was employed for chromosomal integration of *Z. mobilis* genes encoding 10 the ethanol pathway in *E. coli* B (Ohta K *et al.*, (1991) *Appl. Environ. Microbiol.* 57:893-900) and *K. oxytoca* M5A1 (Wood *et al.* (1992) *Appl. Environ. Microbiol.* 58:2103-2110). Typically, circular DNA was transformed into P2 by electroporation using a Bio-Rad Gene Pulser. Next, transformants were selected on solid medium containing tetracycline ( $6 \text{ mg L}^{-1}$ ) and grown on CMC plates to determine levels of 15 glucanase activity.

#### *Glucanase Activity*

- Glucanase activity resulting from expression of *celZ* gene product (*i.e.*, glucanase) under the control of different test promoters was evaluated by staining CMC 20 plates as described in Example 1. This colorimetric assay results in yellow zones indicating glucanase activity and the diameter of the zone was used as a relative measure of *celZ* polypeptide expression. Clones that exhibited the largest zones of yellow color were further evaluated for glucanase activity at  $35^\circ \text{C}$  using carboxymethyl cellulose as the substrate ( $20 \text{ g L}^{-1}$  dissolved in 50 mM citrate buffer, pH 5.2) (Wood *et al.* (1988) 25 *Methods in Enzymology* 160: 87-112). In order to measure the amount of intracellular glucanase, enzymatic activity was released from cultures by treatment with ultra-sound for 4 seconds (Model W-290F cell disruptor, Heat System-Ultrasonics Inc., Plainview, NY). The amount of glucanase activity expressed was measured and is presented here as  $\mu\text{mol}$  of reducing sugar released per min (IU). Reducing sugar was measured as 30 described by Wood (Wood *et al.* (1988) *Methods in Enzymology* 160: 87-112) using a glucose standard.

#### *Substrate Depolymerization*

- To further determine the amount of glucanase activity produced by various host 35 cells, different carbohydrate substrates ( $20 \text{ g L}^{-1}$  suspended in 50 mM citrate buffer, pH 5.2) were incubated with various cell extracts. In one example, test substrates comprising acid-swollen and ball-milled cellulose were prepared as described by Wood

(Wood *et al.* (1988) *Methods in Enzymology* 160: 87-112). A typical polysaccharase extract (*i.e.*, EGZ (glucanase) from *K. oxytoca* SZ6 (pCPP2006)) was prepared by cultivating the host cells at 30°C for 16 h in LB supplemented with sorbitol, a nonreducing sugar. Dilutions of cell-free broth were added to substrates and incubated 5 at 35°C for 16 h. Several drops of chloroform were added to prevent the growth of adventitious contaminants during incubation. Samples were removed before and after incubation to measure reducing sugars by the DNS method (see, Wood *et al.* (1988) *Methods in Enzymology* 160: 87-112). The degree of polymerization (DP) was estimated by dividing the total calculated sugar residues present in the polymer by the 10 number of reducing ends.

#### *Fermentation Conditions*

Fermentations were carried out in 250 ml flasks containing 100 ml of Luria broth supplemented with 50 g L<sup>-1</sup> of carbohydrate. Test carbohydrates were sterilized 15 separately and added after cooling. To minimize substrate changes, acid-swollen cellulose, ball-milled cellulose and xylan were not autoclaved. The antibiotic chloramphenicol (200 mg L<sup>-1</sup>) was added to prevent the growth of contaminating organisms. Flasks were inoculated (10% v/v) with 24-h broth cultures (50 g L<sup>-1</sup> glucose) and incubated at 35° C with agitation (100 rpm) for 24-96 h. To monitor 20 cultures, samples were removed daily to determine the ethanol concentrations by gas chromatography (Dombek *et al.* (1986) *Appl. Environ. Microbiol.* 52:975-981).

#### *Methods for Isolating and Identifying a Surrogate Promoter*

In order to identify random fragments of *Z. mobilis* that would serve as surrogate 25 promoters for the expression of heterologous genes in *Klebsiella* and other host cells, a vector for the efficient cloning of candidate promoters was constructed as described in Example 1 (see also, Ingram *et al.* (1988) *Appl. Environ. Microbiol.* 54:397-404).

Next, *Sau3AI* digested *Z. mobilis* DNA fragments were ligated into the *BamHI* site of pLOI2171 to generate a library of potential promoters. These plasmids were 30 transformed into *E. coli* DH5α for initial screening. Of the 18,000 colonies individually tested on CMC plates, 75 clones produced larger yellow zones than the control (pLOI2173). Plasmids from these 75 clones were then transformed into *K. oxytoca* M5A1, re-tested, and found to express high levels of *celZ* in this second host.

### **Recombinant *Klebsiella* Hosts for Producing Polysaccharase**

The high expressing clones (pLOI2177 to pLOI2194) with the largest zones on CMC plates indicating *celZ* expression were grown in LB broth and assayed for glucanase activity (Table 5).

5

**TABLE 5. Evaluation of promoters for *celZ* expression and secretion in *K. oxytoca* M5A1**

Plasmids <sup>a</sup>	No secretion genes		Secretion genes present (pCPP2006)	
	Total activity (IU L <sup>-1</sup> ) <sup>b</sup>	Secreted activity (IU L <sup>-1</sup> )	Total activity (IU L <sup>-1</sup> )	Secreted activity (IU L <sup>-1</sup> )
PLOI2173	2,450	465	3,190	1,530
PLOI2177	19,700	3,150	32,500	13,300
PLOI2178	15,500	2,320	21,300	11,500
PLOI2179	15,400	2,310	21,400	12,000
PLOI2180	21,400	3,210	30,800	13,600
PLOI2181	15,600	2,490	21,000	11,800
PLOI2182	19,600	3,130	31,100	14,000
PLOI2183	20,700	3,320	32,000	14,000
PLOI2184	15,500	2,480	21,200	11,900
PLOI2185	15,100	2,420	24,600	11,500
PLOI2186	17,000	2,380	25,700	13,400
PLOI2187	15,800	2,210	24,500	12,200
PLOI2188	18,200	2,180	25,600	12,000
PLOI2189	14,800	2,360	27,100	12,700
PLOI2190	16,100	2,410	26,500	12,500
PLOI2191	15,800	2,210	25,000	12,400
PLOI2192	15,100	1,810	24,900	12,500
PLOI2193	16,700	2,010	24,600	12,800
PLOI2194	15,400	2,770	21,500	11,900

<sup>a</sup> pLOI2173 contains the *celZ* gene with the original promoter, all others contain the *celZ* gene with a *Z. mobilis* DNA fragment which serves as a surrogate promoter.

<sup>b</sup> Glucanase (CMCase) activities were determined after 16 h of growth at 30°C.

Activities with these plasmids were up to 8-fold higher than with the control plasmid containing the native *celZ* promoter (pLOI2173). The four plasmids which produced the largest zones (pLOI2177, pLOI2180, pLOI2182 and pLOI2183) also produced the highest total glucanase activities (approximately 30,000 IU L<sup>-1</sup>). One of these plasmids, pLOI2183, was selected for chromosomal integration.

### **Chromosomal Integration of a Polysaccharase Gene**

To stably incorporate a desirable polysaccharase gene into a suitable host cell, e.g., *Klebsiella* P2 strain, a novel vector (pLOI2306) was constructed to facilitate the isolation of a DNA fragment which lacked all replication functions but contained the *celZ* gene with surrogate promoter, a selectable marker, and a homologous DNA fragment for integration (Figure 7). Two *AscI* sites were added to pUC19 by inserting a

linker (GGCGCGCC; SEQ ID NO: 11) into Klenow-treated *Nde*I and *Sap*I sites which flank the polylinker region to produce pLOI2302. A blunt fragment containing the *tet* resistance marker gene from pBR322 (excised with *Eco*RI and *Ava*I, followed by Klenow treatment) was cloned into the *Pst*I site of pLOI2302 (cut with *Pst*I, followed by Klenow treatment) to produce pLOI2303. To this plasmid was ligated a blunt fragment of *K. oxytoca* M5A1 chromosomal DNA (cut with *Eco*RI and made blunt with Klenow treatment) into the *Sma*I site of pLOI2303 to produce (pLOI2305). The *Eco*RI - *Sph*I fragment (Klenow treated) containing the surrogate *Z. mobilis* promoter and *celZ* gene from pLOI2183 was ligated into the *Eco*RI site of pLOI2305 (*Eco*RI, Klenow treatment) to produce pLOI2306. Digestion of pLOI2306 with *Ascl* produced two fragments, the larger of which contained the *celZ* gene with a surrogate promoter, *tet* gene, and chromosomal DNA fragment for homologous recombination. This larger fragment (10 kbp) was purified by agarose gel electrophoresis, circularized by self-ligation, and electroporated into the *Klebsiella* strain P2 and subsequently grown under selection for tetracycline resistance. The resulting 21 tetracycline-resistant colonies were purified and tested on CMC plates for glucanase activity. All were positive with large zones indicating functional expression of the *celZ* gene product.

Clones used to produce the recombinant strains were tested for the presence of unwanted plasmids by transforming DH5 $\alpha$  with plasmid DNA preparations and by gel electrophoresis. No transformants were obtained with 12 clones tested. However, two of these strains were subsequently found to contain large plasmid bands which may contain *celZ* and these were discarded. Both strains with large plasmids contained DNA, which could be sequenced with T7 and M13 primers confirming the presence of multicopy plasmids. The remaining ten strains contain integrated *celZ* genes and could not be sequenced with either primer.

The structural features of the novel vector pLOI2306 are schematically shown in Fig. 8 and the nucleotide sequence of the vector, including various coding regions (*i.e.*, of the genes *celZ*, *bla*, and *tet*), are indicated in SEQ ID NO: 12 of the sequence listing. Nucleotide base pairs 3282-4281, which represent non-coding sequence downstream of the *celZ* gene (obtained from *E. chrysanthemi*), and base pairs 9476-11544 which represent a portion of the non-coding target sequence obtained from *K. oxytoca* M5A1, remain to be sequenced using standard techniques (*e.g.*, as described in Sambrook, J. *et al.*, T. *Molecular Cloning: A Laboratory Manual*. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, (1989); 35 *Current Protocols in Molecular Biology*, eds. Ausubel *et al.*, John Wiley & Sons (1992)). For example, sufficient flanking sequence on either side of the aforementioned unsequenced regions of the pLOI2306 plasmid is provided such that sequencing primers

that correspond to these known sequences can be synthesized and used to carry out standard sequencing reactions using the pLOI2306 plasmid as a template.

Alternatively, it will be understood by the skilled artisan that these unsequenced regions can also be determined even in the absence of the pLOI2306 plasmid for use as 5 a template. For example, the remaining *celZ* sequence can be determined by using the sequence provided herein (e.g., nucleotides 1452-2735 of SEQ ID NO: 12) for synthesizing probes and primers for, respectively, isolating a *celZ* containing clone from a library comprising *E. chrysanthemi* sequences and sequencing the isolated clone using a standard DNA sequencing reaction. Similarly, the remaining target sequence can be 10 determined by using the sequence provided herein (e.g., nucleotides 8426-9475 of SEQ ID NO:12) for synthesizing probes and primers for, respectively, isolating a clone containing target sequence from a library comprising *K. oxytoca* M5A1 *EcoRI* fragments (e.g., of the appropriate size) and sequencing the isolated clone using a standard DNA sequencing reaction (a source of *K. oxytoca* M5A1 would be, e.g., ATCC 15 68564 cured free of any plasmid using standard techniques). The skilled artisan will further recognize that the making of libraries representative of the cDNA or genomic sequences of a bacterium and the isolation of a desired nucleic acid fragment from such a library (e.g., a cDNA or genomic library), are well known in the art and are typically carried out using, e.g., hybridization techniques or the polymerase chain reaction (PCR) 20 and all of these techniques are standard in the art (see, e.g., Sambrook, J. et al., T. *Molecular Cloning: A Laboratory Manual*. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, (1989); *Current Protocols in Molecular Biology*, eds. Ausubel et al., John Wiley & Sons (1992); *Oligonucleotide Synthesis* (M.J. Gait, Ed. 1984); and *PCR Handbook Current Protocols in Nucleic Acid Chemistry*, Beaucage, Ed. John Wiley & Sons (1999) (Editor)). 25

***Heterologous Gene Expression Using a Surrogate Promoter and Integrated or Plasmid-Based Constructs***

The ten integrated strains (SZ1-SZ10) were investigated for glucanase 30 production in LB sorbitol broth (Table 6). All produced 5,000-7,000 IUL<sup>-1</sup> of active enzyme. Although this represents twice the activity expressed from plasmid pLOI2173 containing the native *celZ* promoter, the integrated strains produced only 1/3 the glucanase activity achieved by P2 (pLOI2183) containing the same surrogate *Z. mobilis* promoter (Table 5). The reduction in glucanase expression upon integration 35 may be attributed to a decrease in copy number (i.e., multiple copy plasmid versus a single integrated copy).

### *Secretion of Glucanase EGZ*

*K. oxytoca* contains a native Type II secretion system for pullulanase secretion (Pugsley (1993) *Microbiol. Rev.* 57:50-108), analogous to the secretion system encoded by the *out* genes in *Erwinia chrysanthemi* which secrete pectate lyases and glucanase 5 (EGZ) (Barras *et al.* (1994) *Annu. Rev. Phytopathol.* 32:201-234; He *et al.* (1991) *Proc. Natl. Acad. Sci. USA.* 88: 1079- 1083). Type II secretion systems are typically very specific and function poorly with heterologous proteins (He *et al.* (1991) *Proc. Natl. Acad. Sci. USA.* 88: 1079- 1083; Py *et al.* (1991) *FEMS Microbiol. Lett.* 79:315-322; Sauvonnet *et al.* (1996) *Mol. Microbiol.* 22: 1-7). Thus as expected, recombinant *celZ* 10 was expressed primarily as a cell associated product with either M5A1 (Table 5) or P2 (Table 6) as the host. About 1/4 (12-26%) of the total recombinant EGZ activity was recovered in the broth. With *E. coli* DH5 $\alpha$ , about 8-12% of the total extracellular EGZ was present. Thus the native secretion system in *K. oxytoca* may facilitate partial secretion of recombinant EGZ.

15 To further improve secretion of the desired products, type II secretion genes (*out* genes) from *E. chrysanthemi* EC16 were introduced (e.g., using pCPP2006) to facilitate secretion of the recombinant EGZ from strain P86021 in ethanologenic strains of *K. oxytoca* (Table 5 and Table 6). For most strains containing plasmids with *celZ*, addition of the *out* genes resulted in a 5-fold increase in extracellular EGZ and a 2-fold increase 20 in total glucanase activity. For strains with integrated *celZ*, addition of the *out* genes resulted in a 10-fold increase in extracellular EGZ and a 4-fold increase in total glucanase activity. In both cases, the *out* genes facilitated secretion of approximately half the total glucanase activity. The increase in EGZ activity resulting from addition of the *out* genes may reflect improved folding of the secreted product in both plasmid and 25 integrated *celZ* constructs. The smaller increase observed with the pUC-based derivatives may result from plasmid burden and competition for export machinery during the production of periplasmic  $\beta$ -lactamase from the *bla* gene on this high copy plasmid.

30 Two criteria were used to identify the best integrated strains of P2, growth on solid medium containing high levels of chloramphenicol (a marker for high level expression of the upstream *pdc* and *adhB* genes) and effective secretion of glucanase with the *out* genes. Two recombinant strains were selected for further study, SZ2 and SZ6. Both produced 24,000 IU L<sup>-1</sup> of glucanase activity, equivalent to approximately 5% of the total cellular protein (Py *et al.* (1991) *Protein Engin.* 4:325-333).

### **Substrate Depolymerization**

The substrate depolymerization of the recombinant EGZ was determined to be excellent when applied to a CMC source (Table 7). When applied to acid swollen cellulose, the activity of the glucanase was less than 10% of the activity measured for 5 CMC activity. Little activity was noted when the polysaccharase was applied to Avicel® or xylan. However, when allowed to digest overnight, the EGZ polysaccharase resulted in a measurable reduction in average polymer length for all substrates. CMC and acid-swollen cellulose were depolymerized to an average length of 7 sugar residues. These cellulose polymers of 7 residues are marginally soluble and, ideally, may be 10 further digested for efficient metabolism (Wood *et al.* (1992) *Appl. Environ. Microbiol.* 58:2103-2110). The average chain length of ball-milled cellulose and Avicel® was reduced to 1/3 of the original length while less than a single cut was observed per xylan polymer.

15 **TABLE 6. Comparison of culture growth, glucanase production, and secretion from ethanologenic *K. oxytoca* strains containing integrated celZ**

Strains	Growth on solid medium (600 mg L⁻¹ CM)	Glucanase production and secretion (IU L⁻¹)			
		No secretion system		Adding secretion system (pCPP2006)	
		Total activity	Secreted activity	Total activity	Secreted activity
P2	++++	0	0	0	0
SZ1	++	6,140	1,600	26,100	14,300
SZ2	++++	6,460	1,160	23,700	11,400
SZ3	+++	5,260	1,320	18,400	8,440
SZ4	+++	7,120	1,070	23,200	9,990
SZ5	+	6,000	1,080	29,300	15,500
SZ6	++++	7,620	1,520	24,300	11,900
SZ7	+	6,650	1,330	28,800	15,500
SZ8	+++	7,120	854	28,700	14,900
SZ9	++	7,530	1,130	26,700	12,800
SZ10	+++	4,940	939	17,000	6,600

Glucanase (CMCase) activities were determined after 16 h of growth at 30°C.

20 **TABLE 7. Depolymerization of various substrates by EGZ from cell free broth of strain SZ6 (pCPP2006)**

Substrates	Enzyme activity (IU/L)	Estimated degree of polymerization	
		Before digestion	After digestion
Carboxymethyl cellulose	13,175	224	7
Acid-Swollen cellulose	893	87	7
Ball-milled cellulose	200	97	28
Avicel®	41	104	35
Xylan from oat spelt	157	110	78

Strain SZ6 (pCPP2006) was grown in LB-sorbitol broth for 16 h as a source of secreted EGZ.

**Saccharification and Fermentation Ability of a Biocatalyst**

To be useful, addition of *celZ* and *out* genes to strain P2 must not reduce the fermentative ability of the resulting biocatalyst. A comparison was made using glucose and cellobiose (Table 8). All strains were equivalent in their ability to ferment these sugars indicating a lack of detrimental effects from the integration of *celZ* or addition of pCPP2006. These strains were also examined for their ability to convert acid-swollen cellulose directly into ethanol. The most active construct SZ6 (pCPP2006) produced a small amount of ethanol ( $3.9 \text{ g L}^{-1}$ ) from amorphous cellulose. Approximately  $1.5 \text{ g L}^{-1}$  ethanol was present initially at the time of inoculation for all strains. This decreased with time to zero for all strains except SZ6 (pCPP2006). Thus the production of  $3.9 \text{ g L}^{-1}$  ethanol observed with SZ6 (pCPP2006) may represent an underestimate of total ethanol production. However, at best, this represents conversion of only a fraction of the polymer present. It is likely that low levels of glucose, cellobiose, and celotriose were produced by EGZ hydrolysis of acid swollen cellulose and fermented. These compounds can be metabolized by the native phosphoenolpyruvate-dependent phosphotransferase system in *K. oxytoca* (Ohta K et al., (1991) *Appl. Environ. Microbiol.* 57:893-900; Wood et al. (1992) *Appl. Environ. Microbiol.* 58:2103-2110).

**TABLE 8. Ethanol production by strain SZ6 containing *out* genes (pCPP2006) and integrated *ce/Z* using various substrates ( $50 \text{ g L}^{-1}$ )**

Strains	Ethanol production ( $\text{g L}^{-1}$ )		
	Glucose	Cellobiose	Acid-swollen cellulose
P2	22.9	22.7	0
P2 (pCPP2006)	22.6	21.3	0
SZ6	21.5	19.7	0
SZ6 (pCPP2006)	22.7	21.2	3.9

Initial ethanol concentrations at the time of inoculation were approximately  $1.5 \text{ g L}^{-1}$  for all cultures. With acid swollen cellulose as a substrate, these levels declined to 0 after 72 h of incubation for all strains except SZ6 (pCP206).

**EXAMPLE 3**

**Synergistic Hydrolysis of Carboxymethyl Cellulose and Acid-swollen Cellulose by Two Endoglucanases (EGZ and EGY) from *Erwinia chrysanthemi***

This example describes production of the endoglucanases EGY and EGZ by recombinant *E. coli* and the synergistic hydrolysis of carboxymethyl cellulose (CMC) and acid-swollen cellulose by these endoglucanases.

Throughout this example, the following materials and methods are used unless otherwise stated.

## **Materials and Methods**

### *Bacteria, plasmids and culture conditions*

Bacterial strains and plasmids used in this study are listed in Table 9.

5 **TABLE 9. Strains and plasmids used**

<b>Strains / Plasmids</b>	<b>Descriptions</b>	<b>References / Sources</b>
<b>Strains</b>		
<i>Escherichia coli</i>		
DH5α	<i>lacZ M15 recA</i>	Bethesda Research Laboratory
B	Prototrophic	ATCC11303
HB101	<i>recA lacY recA</i>	ATCC37159
TOP10F'	This strain expresses the <i>lac</i> repressor ( <i>lacI<sup>q</sup></i> gene) from an F episome	Invitrogen
<b>Plasmids</b>		
pCR2.1-TOPO	TOPO™ TA Cloning vector, Ap <sup>r</sup> , Km <sup>r</sup>	Invitrogen
pRK2013	Km <sup>r</sup> mobilizing helper plasmid ( <i>mob</i> <sup>+</sup> )	ATCC
pCPP2006	Sp <sup>r</sup> , ca. 40 kbp plasmid carrying the complete <i>out</i> genes from <i>E. chrysanthemi</i> EC16	He, et al. (1991) <i>Proc. Natl. Acad. Sci. USA</i> 88:1079-1083.
pLOI1620	Ap <sup>r</sup> , <i>celZ</i> gene and its native promoter from <i>E. chrysanthemi</i> P86021	Beall, et al. (1993) <i>J. Indust. Microbiol.</i> 11:151-155.
pMH18	Ap <sup>r</sup> , <i>celY</i> gene and its native promoter from <i>E. chrysanthemi</i> 3937	Guiseppi, et al (1991) <i>Gene</i> 106:109-114.
pLOI2311	<i>celY</i> gene (without native promoter), cloned into pCR2.1-TOPO vector and oriented for expression from the <i>lac</i> promoter	See text

*Escherichia coli* DH5α and TOP10F' were used as hosts for plasmid constructions.

The *celZ* gene was cloned from *E. chrysanthemi* P86021 (Beall, et al. (1993) *J. Indust.*

- 10 *Microbiol.* 11:151-155). The *celY* gene was cloned by Guiseppi et al. ((1991) *Gene* 106:109-114) from *E. chrysanthemi* 3937. The *out* genes were cloned by He et al. ((1991) *Proc. Acad. Sci. USA* 88:1079-1083) from *E. chrysanthemi* EC16.

*E. coli* cultures were grown at 37°C in Luria-Bertani broth (LB) containing per liter: 10 g Difco tryptone, 5 g Difco® yeast extract, and 5 g sodium chloride or on solid

- 15 LB medium containing agar (1.5 %). Clones were screened for endoglucanase production using the Congo Red method (Wood, et al. (1989) *Biochem. J.* 260:37-43). Indicator plates were prepared by supplementing LB agar with low viscosity CMC (0.3%). Ampicillin (50 µg/ml), kanamycin (50 µg/ml) and spectinomycin (100 µg/ml) were added as appropriate for selection.

### *Genetic methods*

- Standard methods were used for plasmid construction and analyses (Ausubel, *et al.* (1987) *Current Protocols in Molecular Biology*. New York: John Wiley and Sons, Inc). The coding region for *celY* was amplified by the polymerase chain reaction using 5 pMH18 as the template with the following primer pairs: N-terminus 5'CTGTTCCGTTACCAACAC3' (SEQ ID NO:13)', C-terminus 5'GTGAATGGGATCACGAGT3' (SEQ ID NO:14). The *E. chrysanthemi* out genes (pCPP2006) were transferred by conjugation using pRK2013 for mobilization (Zhou, *et al.* (1999) *B. Appl. Environ. Microbiol.* 65:2439-2445). DNA was sequenced by the 10 dideoxy method using a LI-COR Model 4000-L DNA sequencer and fluorescent primers.

### *Enzyme assay*

- Endoglucanase activity was determined *in vitro* using CMC as a substrate. 15 Appropriate dilutions of cell-free culture broth (extracellular activity) or broth containing cells that had been disrupted by ultrasound (total activity) were assayed at 35°C in 50 mM citrate buffer (pH 5.2) containing low viscosity CMC (20 g per liter). Reactions were terminated by heating in a boiling water bath for 10 min. Reducing sugars were measured using 3,5-dinitrosalicylic acid reagent with glucose as a standard 20 (Wood, *et al.* (1988) *Methods Enzymology* 160:87-112). Enzyme activity (CMCase) is expressed as μmol reducing sugar released per min (IU). Results are an average of two or more determinations.

### *Synergism*

- 25 Stationary phase cultures of DH5α(pLOI1620 + pCPP2006) and DH5α(pLOI2311) were sonicated and centrifuged as described in Zhou, *et al.* (1999), *supra*, as a source of EGZ and EGY, respectively. These were diluted as necessary to provide equal CMCase activities. Mixtures of EGZ and EGY were tested for synergy at 35°C in 50 mM citrate buffer (pH 5.2) containing CMC (20 g/L) or acid swollen 30 cellulose (20 g per liter). For tests with Avicel® (20 g per liter), enzyme preparations were mixed without prior dilution. Hydrolyzed samples of acid-swollen cellulose and Avicel® were centrifuged (10,000 x g, 5 min) to remove insoluble material prior to the determination of reducing sugars.

The effect of sequential additions of EGZ and EGY was also investigated.

- 35 Substrates were hydrolyzed with a single enzyme for 4 hours and then inactivated by boiling for 20 minutes. After cooling, the second enzyme was added and incubated for an additional 4 hours. Control experiments were conducted with both enzymes together

GENETIC ENGINEERING  
INSTITUTE OF BIOTECHNOLOGY

(4 hours) and with each enzyme alone (4 hours). Samples were analyzed for reducing sugar. In some cases, products were also analyzed by thin layer chromatography.

The degree of synergism for enzyme mixtures was calculated as the observed activity divided by the sum of predicted contributions from EGY alone and EGZ alone

5 (Riedel, *et al.* (1997) *FEMS Microbiol. Lett.* 147:239-243).

#### *Hydrolysis of cellooligosaccharides*

Hydrolysis products from cellobiose, cellotriose, cellotetraose, cellopentaose, acid-swollen cellulose (Wood, *et al.* (1988), *supra*) and Avicel® were analyzed by thin 10 layer chromatography. For these analyses, 15 µl of 1% substrate was mixed with 45 µl of crude enzyme (0.07 IU), incubated at 35°C for 2 hours and terminated by heating in a boiling water bath. Hydrolysates were spotted on Whatman 250 µm Silica-gel 150A plates and developed for approximately 4 hours using the solvent system described by Kim, (1995) *Appl. Environ. Microbiol.* 61:959-965). By volume, this solvent contained 15 6 parts chloroform, 7 parts acetic acid, and 1 part water. Sugars were visualized by spraying of 6.5 mM N-(1-naphthyl) ethylenediamine dihydrochloride and heating at 100°C for approximately 10 min (Bounias, (1980) *Anal. Biochem.* 106:291-295).

#### *Materials and chemicals*

20 Tryptone and yeast extract were products of Difco (Detroit, Michigan). Antibiotics, low viscosity CMC, cellobiose, cellotriose, and cellotetraose were obtained from the Sigma Chemical Company (St. Louis, Missouri). Cellopentaose was obtained from V-Lab (Covington, Louisiana). Avicel® was purchased from Fluka Chemika (Buchs, Switzerland).

25

#### *Production of EGY and EGZ by recombinant *E. coli**

Low levels of EGY activity were produced by native *E. chrysanthemi* 3937 and by recombinant *E. coli* harboring plasmid pMH18 (Boyer, *et al.* ((1987) *Eur. J. Biochem.* 162:311-316; Guiseppi, *et al.*, *supra*). Poor expression from the high copy 30 plasmid in *E. coli* was attributed to promoter function and a putative requirement for a *celY* activator protein (Guiseppi, *et al.*, *supra*). A new clone was constructed to produce higher levels of EGY for our investigations of synergy. The EGY coding region (without promoter) was amplified using the polymerase chain reaction and cloned behind the *lac* promoter in pCR2.1-TOPO. The resulting plasmid, pLOI2311, was 35 strongly positive on CMCase indicator plates. Replacement of the native promoter with the *lac* promoter increased *celY* expression by approximately 10-fold, from 165 IU/L to 1800 IU/L (See Table 10, below).

**TABLE 10. Effect of *E. chrysanthemi* *out* genes on the expression and secretion of celY and celZ in *E. coli* DH5 $\alpha$**

5

Enzyme expressed	Promoter	Growth (hr)	No <i>out</i> genes			<i>out</i> genes present (pCPP2006)		
			Extracellular CMCase <sup>a</sup> (IU per liter)	Total CMCase (IU per liter)	Apparent secretion (%)	Extracellular CMCase <sup>a</sup> (IU per liter)	Total CMCase (IU per liter)	Apparent secretion (%)
EGY	Native promoter (pMH18)	24	136	165	82	136	180	76
	<i>lac</i> promoter (pLOI2311)	8	208	266	78	nd <sup>b</sup>	Nd	nd
		16	1,420	1,590	90	nd	Nd	nd
		24	1,650	1,800	90	1,360	1,510	90
EGZ	Native plus <i>lac</i> promoter (pLOI1620)	8	130	1,320	10	6,710	7,4600	90
		16	1,200	9,030	13	13,400	19,700	68
		24	1,800	12,500	14	23,600	36,800	64

<sup>a</sup> Secreted or released CMCase activity in the culture supernatant.

<sup>b</sup> Abbreviation: nd, not determined.

10

Approximately 90% of EGY activity was found in the extracellular milieu.

Expression of celZ was included for comparison (See Table 10). High levels of EGZ were produced by *E. coli* harboring plasmid pLOI1620. Extracellular EGZ and total EGZ activity were further increased by addition of the *E. chrysanthemi* *out* genes

15 (pCPP2006) as reported previously (Zhou, et al. (1999) *Appl. Environ. Microbiol.* 65:2439-2445). Unlike EGZ, however, EGY activity was not affected by the presence of *out* genes. Maximal EGY and EGZ activities were obtained from 24-hour cultures. The supernatants from disrupted cultures of DH5 $\alpha$  containing pLOI2311 or pLOI1620 and pCPP2006 (*out* genes) were used as a source of EGY and EGZ, respectively, for 20 further investigations.

#### **Synergistic action of EGY and EGZ with CMC as a substrate**

Initial studies examining the combined actions of EGY and EGZ were conducted with CMC (20 g per liter) for a single incubation time (Figure 9). Disrupted cell 25 preparations containing EGY and EGZ were each diluted to equal activities (CMCase) and combined in different proportions to maintain a constant sum of individual activities. EGY and EGZ were tested individually as controls. All mixtures of EGY and EGZ were significantly higher than either enzyme assayed alone indicating a synergistic interaction. The synergistic effect increased with the proportion of EGZ. Maximal 30 synergy (1.42) was observed with a ratio of EGZ to EGY activities of 9 to 1 and 19 to 1.

Further experiments examined the effect of incubation time using CMC as the substrate and an activity ratio of 9 to 1 for EGZ and EGY, respectively (Figure 9). EGZ and EGY alone were included as controls. The synergistic effect of combining EGZ and EGY was clearly evident as an increase in the rate and extent of hydrolysis. Calculated synergy increased with incubation time. At the end of the incubation (4 hours), the concentration of reducing sugars was 1.8-fold higher in the mixed enzyme preparation than predicted by the arithmetic sum of individual EGZ and EGY activities, *i.e.*, synergistic activity.

10 ***Effect of substrate (CMC) concentration on synergy***

The effect of substrate concentration on the synergy between EGZ and EGY was also studied (See Table 11, below). Increasing the CMC concentration from 2.5 g per liter to 20 g per liter increased the observed synergy from 1.12 to 1.89. Based on the specific activities of EGZ and EGY and a maximal synergism of 1.89, the enzyme turnover rate for the combination was 8-fold that of purified EGY alone, and 1.5-fold that of purified EGZ alone.

**TABLE 11. Effect of substrate concentration on synergy**

CMC Substrate (g/L)	Reducing sugar released ( $\mu$ mole/ml) <sup>a</sup>			Synergy <sup>a,c</sup>
	EGZ (10) <sup>b</sup>	EGY(10) <sup>b</sup>	EGZ(9) + EGY(1) <sup>b</sup>	
20	3.98 $\pm$ 0.04	3.83 $\pm$ 0.04	7.51 $\pm$ 0.07	1.89 $\pm$ 0.02
10	4.53 $\pm$ 0.01	2.91 $\pm$ 0.07	5.38 $\pm$ 0.04	1.25 $\pm$ 0.01
5.0	2.87 $\pm$ 0.01	1.18 $\pm$ 0.04	2.92 $\pm$ 0.04	1.08 $\pm$ 0.02
2.5	1.42 $\pm$ 0.01	0.50 $\pm$ 0.04	1.49 $\pm$ 0.01	1.12 $\pm$ 0.01

20

<sup>a</sup> Average  $\pm$  standard deviation.

<sup>b</sup> EGZ and EGY were diluted to equal CMCase activities. Reactions (0.15 IU/ml) contained 9 parts of EGZ and 1 part of EGY. As controls, EGZ (0.15 IU/ml) and EGY (0.15 IU/ml) were each tested individually.

25

<sup>c</sup> Synergy was calculated as the observed activity divided by the sum of predicted contributions from EGY alone (10%) plus EGZ alone (90%).

EGY was more sensitive to substrate concentration than EGZ. Increasing the CMC concentration resulted in an 8-fold increase in reducing sugar products with EGY 30 but only a 3-fold increase with EGZ. Based on a double reciprocal plot of the data in Table 11, apparent Km values of 104, 12, and 38 g per liter were estimated for EGY, EGZ and the combination of both enzymes (9 parts EGZ + 1 part EGY), respectively. The higher apparent Km for EGY is consistent with a requirement for longer substrate molecules.

- The extent of CMC hydrolysis was also examined by determining the approximate size of hydrolysis products. CMC (1.25 g per liter) was incubated (4 hours, 0.75 IU CMCase/ml) with EGY, EGZ, and a combination of both enzymes (9 parts EGZ plus 1 part EGY). Chain length was estimated based on the reducing sugar assay before 5 (250 glucosyl units) and after incubation. The average chain length was substantially reduced by all three enzyme preparations. EGZ was more effective in reducing chain length than EGY, 3.6 versus 10.7 glucosyl residues, respectively. The combination of both enzymes resulted in a synergistic action. Simultaneous hydrolysis with both enzymes reduced the average size of the hydrolysis products to 2.3 glucosyl residues,
- 10 36% lower than EGZ alone and 79% lower than EGY alone. These results confirm that EGZ readily hydrolyzes both large CMC polymers and smaller saccharides. The action of EGY appears more limited, primarily hydrolyzing large polymers with greater than 10 glucosyl units.
- 15 ***Sequential and simultaneous hydrolysis of CMC with EGZ and EGY***
- The mechanism of synergistic action between EGZ and EGY was further investigated by comparing the effects of sequential hydrolysis with individual endoglucanases to that of simultaneous hydrolysis by a mixture of both enzymes (See Table 12, below). Again synergy was observed for the simultaneous actions of both 20 enzymes. No synergy was observed for the sequential hydrolysis of CMC when EGZ was used as the first enzyme, followed by digestion with EGY (after heat-inactivation of EGZ). In contrast, full synergy was retained when CMC was first treated with EGY, followed by EGZ (after heat-inactivation of EGY). These results indicate that synergy can be achieved by the independent activities of EGY and EGZ. Enzymatic 25 modification of the substrate by EGY increased the rate and extent of subsequent hydrolysis by EGZ. These results provide further evidence that EGY and EGZ function quite differently. EGY appears to primarily reduce the chain length of large polymers while EGZ appears to act more randomly, hydrolyzing both large and small substrate molecules.

RECORDED  
SEARCHED  
INDEXED  
SERIALIZED  
FILED

**TABLE 12. Sequential and simultaneous hydrolysis of CMC by EGZ and EGY**

Enzyme (relative proportion) <sup>a</sup>	Measured reducing sugar released ( $\mu\text{mole/ml}$ ) <sup>b</sup>	Predicted activity from the arithmetic sum of EGY and EGZ ( $\mu\text{mole/ml}$ ) <sup>c</sup>	Synergy <sup>b,d</sup>
EGZ (10) + EGY (0)	4.65 ± 0.08	4.65	1.00 ± 0.02
EGZ (0) + EGY (10)	4.14 ± 0.04	4.14	1.00 ± 0.01
EGZ (9) + EGY (1). (simultaneously)	8.28 ± 0.08	4.60	1.80 ± 0.02
EGZ (9), then EGY (1). (sequential)	4.86 ± 0.23	4.60	1.06 ± 0.05
EGY (1), then EGZ (9). (sequential)	8.75 ± 0.14	4.60	1.90 ± 0.03

<sup>a</sup> EGZ and EGY were diluted to equal CMCase activities. Both simultaneous and sequential hydrolysis reactions (0.15 IU/ml) were investigated using 9 parts of EGZ and 1 part of EGY. In the sequential hydrolysis experiments, the first enzyme was incubated with substrate for 4 hours and inactivated by boiling for 20 min. After cooling, the second enzyme was added and incubated for an additional 4 hours. All reactions were terminated by boiling.

<sup>b</sup> Average ± standard deviation (3 experiments).

<sup>c</sup> Calculated sum of individual EGY and EGZ activities.

<sup>d</sup> Synergy was calculated as the observed activity divided by the sum of predicted contributions from EGY alone (10%) plus EGZ alone (90%).

#### *Synergistic action on acid-swollen and crystalline cellulose*

Potential synergy was investigated using acid-swollen cellulose as the substrate and a 9 to 1 ratio of EGZ:EGY based on CMCase activities (Figure 9). Since the activities of EGZ and EGY with acid-swollen cellulose are lower than those with CMC (Boyer, *et al.* (1987) *Eur. J. Biochem.* 162:311-316), enzyme loading (1.5 IU) and incubation times were increased. When assayed individually with acid-swollen cellulose, EGY was approximately 1/3 as active as EGZ. However, the combination of these two enzymes was significantly more active than the predicted arithmetic sum of individual activities at all time points. The degree of synergy was essentially constant (1.36 ± 0.17) during the 36 hour period of incubation.

The hydrolysis products from acid-swollen cellulose (6 hours) were analyzed by thin layer chromatography (Figure 10). No soluble saccharides were observed after incubation with EGY alone. Cellobiose and cellotriose were the primary products from hydrolysis with EGZ alone and a combination of EGY and EGZ. With the combination of both enzymes, higher product levels were evident as darker and larger spots confirming a synergistic action.

The synergistic action of EGZ and EGY was also investigated with Avicel® (Figure 10), a highly crystalline cellulose. Small amounts of cellobiose and cellotriose were observed as hydrolysis products with EGZ alone and with the mixture of EGY and EGZ. Due to low activity with Avicel®, high loadings (10 µl) were required on thin

layer plates to visualize products. Note that this additional salt increased the relative migration of oligosaccharide products in comparison to the standards. No cellooligosaccharide spots were observed with EGY alone. Again synergism was evident with the combination of EGY and EGZ. Larger and more intense spots were  
5 observed corresponding to cellobiose, cellotriose, and cellotetraose with the combined activities than with EGZ alone. The low activity with Avicel® as a substrate and the relatively low levels of products are consistent with the hydrolysis of the amorphous rather than the crystalline regions of Avicel®. These results indicate that the synergistic action of EGZ and EGY is not limited to a model substrate such as CMC. Synergistic  
10 hydrolysis was also observed for acid-swollen cellulose and the amorphous regions of Avicel®.

#### ***Hydrolysis of cellooligosaccharides***

The substrate specificities of EGZ and EGY were further investigated using  
15 soluble cellooligosaccharides (cellobiose, cellotriose, cellotetraose, and cellopentaose). Hydrolysis products were analyzed by thin layer chromatography (Figure 11). Cellobiose was not hydrolyzed by EGY, EGZ or a combination of both enzymes. None of the cellooligosaccharides was hydrolyzed by EGY alone (Figure 11, Panel B). In contrast, EGZ hydrolyzed cellotetraose and cellopentaose but not cellotriose (Figure 11,  
20 Panel C). EGZ hydrolysis products from cellotetraose were primarily cellobiose with lesser amounts of cellotriose and glucose. With cellopentaose as the substrate, EGZ produced approximately equal amounts of cellobiose and cellotriose indicating a preferential attack on the second or third glycosidic bond. This was further confirmed by examining samples at various times during the incubation of cellopentaose with EGZ  
25 (Figure 11, Panel D). Cellobiose and cellotriose progressively accumulated during incubation with a corresponding reduction in cellopentaose. Thus in contrast to the requirement for large substrates by EGY, EGZ hydrolyzes soluble cellooligosaccharides containing four or more glucosyl units.

30

#### **EXAMPLE 4**

##### **Integration, Expression, and Extracellular Secretion of *Erwinia chrysanthemi* Endoglucanase EGY (*celY*) and EGZ (*celY*) in Ethanologenic *Klebsiella oxytoca* P2**

In this example, the functional integration of both *celY* and *celZ* from *E. chrysanthemi* into the chromosome of *K. oxytoca* P2, is described. Also described is the  
35 synergism between recombinant EGY and EGZ and fungal cellulase (Spezyme CE®) during the fermentation of cellulose to ethanol using simultaneous saccharification and fermentation.

Throughout this example, the following materials and methods are used unless otherwise stated.

### **Materials and Methods**

#### **5 Bacteria, plasmids and culture conditions**

Strains and plasmids used in this example are listed in Table 13 below.

**TABLE 13. Strains and Plasmids**

<b>Strains / plasmids</b>	<b>Description</b>	<b>Sources / references</b>
<b><i>E. coli</i> strains</b>		
DH5α	<i>lacZ M15 recA</i>	Bethesda Research Laboratory
TOP10F'	<i>hsdR mcrA lacZΔM15 endA recA; F' tet lacI</i>	Invitrogen
HB101	<i>recA lacY</i>	ATCC37159
S17-1	<i>thi pro recA hsdR RP4-2-tet::Mu aphaA:Tn7 λpir</i>	De Lorenzo, <i>et al.</i> (1990) <i>J. Bacteriol.</i> 172:6568-6572.
<b><i>Z. mobilis</i> strain</b>		
CP4	Prototrophic	Ingram, <i>et al.</i> (1999) <i>Biotechnol. Prog.</i> 15:855-866.
<b><i>K. oxytoca</i> strains</b>		
M5A1	Prototrophic	Wood, <i>et al.</i> (1992) <i>Appl. Environ. Microbiol.</i> 58:2103-2110.
P2	<i>pfl::pdc adhB cat</i>	Wood, <i>et al.</i> (1992) <i>Appl. Environ. Microbiol.</i> 58:2103-2110.
SZ6	<i>pfl::pdc adhB cat; integrated celZ tet</i>	
SZ12	<i>pfl::pdc adhB cat; integrated celZ celY kan</i>	See text
SZ21	<i>pfl::pdc adhB cat; integrated celZ celY</i>	See text
SZ22	<i>pfl::pdc adhB cat; integrated celY celZ::aac</i>	See text
<b>Plasmids</b>		
pUC18	<i>bla</i> cloning vector	New England Biolabs
pUC19	<i>bla</i> cloning vector	New England Biolabs
PCR2.1-TOPO	TA cloning vector, <i>bla kan</i>	Invitrogen
pMH18	<i>bla celY</i> from <i>E. chrysanthemi</i> 3937	Guiseppi, <i>et al.</i> (1991) <i>Gene</i> 106:109-114.
pHPΩ245aac	<i>bla aac</i> source of apramycin gene	Blondelet-Rouault, <i>et al.</i> (1997) <i>Gene</i> 190:315-317.
pBR322	<i>bla tet</i> cloning vector	New England Biolabs
PRK2013	<i>kan</i> , mobilizing plasmid	ATCC

**TABLE 13. Strains and Plasmids (Continued)**

pCPP2006	<i>spm</i> , ca. 40 kbp fragment containing <i>out</i> genes from <i>E. chrysanthemi</i> EC16	He, <i>et al.</i> (1991) <i>Proc. Natl. Acad. Sci. USA</i> 88:1079-1083.
pFT-A	<i>bla</i> low copy vector containing <i>flp</i> recombinase gene and temperature conditional pSC101 replicon	Martinez-Morales, <i>et al.</i> (1999) <i>J. Bacteriol.</i> 181:7143-7148.
pLOI2224	<i>kan</i> integration vector containing conditional R6K replicon and two FRT sites	Martinez-Morales, <i>et al.</i> (1999) <i>J. Bacteriol.</i> 181:7143-7148.
pLOI2307	<i>bla</i> containing <i>celZ</i> gene and a surrogate promoter from <i>Z. mobilis</i> DNA	Zhou, <i>et al.</i> (1999) <i>Appl. Environ. Microbiol.</i> 65:2439-2445.
pLOI2311	PCR fragment containing <i>celY</i> gene cloned into pCR2.1-TOPO, expressed from <i>lac</i> promoter	Zhou, <i>et al.</i> (2000) <i>J. Bacteriol.</i> 182:5676-5682.
pLOI2302	pUC19 containing <i>AscI</i> linkers inserted into blunt <i>NdeI</i> and <i>SapI</i> sites	Zhou, <i>et al.</i> (1999) <i>J. Indust. Microbiol. Biotechnol.</i> 22:600-607.
pLOI2316	pUC 18 containing the <i>celY</i> gene on a Klenow-treated <i>EcoRI</i> fragment from pLOI2311 inserted into a blunt <i>HincII</i> site, expressed from the <i>lac</i> promoter	See text
pLOI2317	<i>EcoRI-HindIII</i> fragment from pLOI2316 inserted into the corresponding sites of pLOI2302	See text
pLOI2318	<i>Sau3A</i> 1 fragment of <i>Z. mobilis</i> DNA fragment exhibiting promoter activity inserted into the <i>BamHI</i> site of pLOI2317	See text
pLOI2319	<i>Sau3A</i> 1 fragment of <i>Z. mobilis</i> DNA exhibiting promoter activity inserted into the <i>BamHI</i> site of pLOI2317	See text
pLOI2320	<i>Sau3A</i> 1 fragment of <i>Z. mobilis</i> DNA exhibiting promoter activity inserted into the <i>BamHI</i> site of pLOI2317	See text
pLOI2323	<i>Sau3A</i> 1 fragment of <i>Z. mobilis</i> DNA exhibiting promoter activity inserted into the <i>BamHI</i> site of pLOI2317	See text
pLOI2342	<i>Sau3A</i> 1 fragment of <i>Z. mobilis</i> DNA exhibiting promoter activity inserted into the <i>BamHI</i> site of pLOI2317	See text
pLOI2348	Random <i>EcoRI</i> fragment of <i>K. oxytoca</i> M5A1 DNA cloned into <i>EcoRI</i> site of pLOI2323	See text
pLOI2349	<i>EcoRI</i> linker inserted into the Klenow-treated <i>SphI</i> site of pLOI2307	See text
pLOI2350	<i>EcoRI</i> fragment ( <i>celZ</i> and surrogate promoter) from pLOI2349 inserted into the <i>EcoRI</i> site of pLOI2224	See text
pLOI2352	<i>AscI</i> fragment ( <i>K. oxytoca</i> fragment, <i>Z. mobilis</i> promoter fragment and <i>celY</i> ) from pLOI2348 inserted into the <i>AscI</i> site of pLOI2350	See text

**TABLE 13. Strains and Plasmids (Continued)**

pLOI2353	<i>EcoRI-Aval</i> fragment ( <i>tet</i> gene) from pBR322 inserted into the <i>Clal</i> site of pFT-A.	See text
pLOI2354	pUC19 derivative in which the multiple cloning sites from <i>Hind</i> III to <i>Sma</i> I were deleted by digestion, Klenow-treatment, and self-ligation	See text
pLOI2355	<i>EcoRI</i> fragment ( <i>celZ</i> gene) from pLOI2349 inserted into the <i>EcoRI</i> site of pLOI2354.	See text
pLOI2356	<i>Sma</i> I fragment containing the apramycin resistance gene ( <i>aac</i> gene) from pHPO45aac inserted into the T4 polymerase-treated <i>Pst</i> I site of pLOI2355, disrupting the <i>celZ</i> gene	See text
pLOI2357	<i>EcoRI</i> fragment ( <i>aac</i> and disrupted <i>celZ</i> ) inserted into the <i>EcoRI</i> site of pLOI2224	See text
pLOI2358	Subclone of pLOI2323 in which the internal <i>Pst</i> I fragment was deleted, used for sequencing	See text
pLOI2359	Subclone of pLOI2323 in which the <i>Clal-Hind</i> III fragment was deleted, used for sequencing	See text

5      *Escherichia coli* DH5 $\alpha$  and TOPO10F' were used as hosts during plasmid constructions. The *celZ* gene, *celY* gene, and *out* genes were cloned as described in Example 3.

10     *E. coli* cultures were grown at 37°C in Luria-Bertani broth (LB) containing per liter: 10 g Difco® tryptone, 5 g Difco® yeast extract, and 5 g sodium chloride or on solid LB medium containing agar (1.5%). Sugar was always included in broth (5% glucose or sorbitol) and solid media (2% glucose) used for the growth of ethanologenic strains. Clones were screened for endoglucanase production using the Congo Red method (Wood *et al.* (1988) *Methods Enzymology* 160:87-112). Endoglucanase indicator plates were prepared by supplementing LB agar with 0.3% low viscosity carboxy methyl cellulose (CMC). Ampicillin (50 µg/ml), apramycin (100 µg/ml), kanamycin (50 µg/ml), chloramphenicol (40 µg/ml) and spectinomycin (100 µg/ml) were used for selection. Ethanologenic strains of *K. oxytoca* were maintained at 30°C on solid LB medium containing glucose (2%) and chloramphenicol (600 µg/ml).

#### *Genetic methods*

20     Standard methods were used for plasmid construction, analyses, and sequencing. The ribosome-binding site and promoterless coding region of *celY* were amplified by the polymerase chain reaction using pMH18 as the template with the following primer pairs: N-terminus 5'CTGTTCCGTTACCAACAC3' (SEQ ID NO:13), C-terminus 5'GTGAATGGGATCACGAGT3' (SEQ ID NO:14). The *E. chrysanthemi* *out* genes (pCPP2006) were transferred by conjugation using pRK2013 for mobilization.

Constructions were confirmed by sequencing using the dideoxy method and a LI-COR Model 4000-L DNA sequencer with fluorescent primers. The *E. chrysanthemi celY* and *celZ* genes were introduced into *K. oxytoca* P2 by electroporation using a Bio-Rad Gene Pulser®. Recombinants were selected on solid medium containing kanamycin (50 mg/liter) as described in Martinez-Morales, *et al.* (1999) *J. Bacteriology* 181:7143-7148, and Zhou, *et al.* (1999) *Indust. Microbiol. Biotechnol.* 22:600-607.

#### *Primer extension analysis*

Promoter regions were identified by mapping the transcriptional start sites using 10 aIRD41-labeled primer fluorescent primers within the coding regions: 5'- ACCATCAGCATCAACGCCAACAAACG-3' (SEQ ID NO: 15) for *celY* and 5'- GACTGGATGGTTATCCGAATAAGAGAGAGG-3' (SEQ ID NO: 16) for *celZ*. Extension products were dissolved in loading buffer and compared to parallel dideoxy sequences using the LI-COR Model 4000-L DNA sequencer (Lincoln, Nebraska).

15

#### *Enzyme assay*

Endoglucanase activity was determined as described in Example 3.

#### *Fermentation*

20 Simultaneous saccharification and fermentation (SSF) tests were conducted in unbaffled, 500-ml flasks containing a 200 ml of broth. Flasks were fitted with a rubber stopper and vented with an 18 gauge needle. Fermentations were conducted at 35°C (120 rpm) in LB medium containing 10% Sigmacell 50 (crystalline cellulose). Inocula were grown for 12 hours in LB containing 5% glucose. Cells were harvested by 25 centrifugation and resuspended in LB. Each flask was inoculated to provide an initial density of 16 mg of cells (dry weight).

#### *Materials and chemicals*

Tryptone and yeast extract were products of Difco (Detroit, Michigan). 30 Antibiotics, low viscosity CMC, and Sigmacell 50® were obtained from the Sigma Chemical Company (St. Louis, Missouri). The I RD41-labeled fluorescent primers were purchased from LI-COR, Inc. (Lincoln, Nebraska).

#### *Construction of a promoter-probe vector for celY*

35 The *celY* gene from *E. chrysanthemi* is poorly expressed from its native promoter in *E. coli* (See Guiseppe, *et al* (1991) *Gene* 106:109-114). Accordingly, to increase expression, a promoter-probe vector was constructed as follows using *celY* as

the reporter (See Figure 13). A promoterless *celY* coding region with ribosomal-binding site (1.2 kbp) was amplified by PCR using pMH18 as the template and randomly inserted into the topoisomerase vector, PCR2.1-TOPO. Functional expression of *celY* was confirmed using endoglucanase indicator plates. A clone oriented to express *celY* from the *lac* promoter was selected and designated pLOI2311 (5.2 kbp). An *Eco*RI fragment containing the promoterless *celY* gene was isolated from pLOI2311. The ends of this fragment were blunted using Klenow polymerase prior to ligation into the *Hinc*II site of pUC18. A clone oriented to express *celY* from the *lac* promoter was selected (3.9 kbp) and expression confirmed using endoglucanase indicator plates (pLOI2316). The promoterless *celY* gene was isolated from pLOI2316 as a 1.2 kbp fragment using *Eco*RI and *Hind*III and inserted into the corresponding sites of pLOI2302 (pUC19 derivative) to reverse the direction of the *celY* gene. As expected, the resulting construct DH5 $\alpha$ (pLOI2317) was inactive on endoglucanase indicator plates due to the lack of a promoter. To facilitate the insertion of DNA fragments containing promoter regions, plasmid pLOI2317 (3.9 kbp) contains a *Bam*HI site in the polylinker region, immediately upstream from the *celY* gene (See Figure 13).

***Construction of plasmids with increased expression of celY in E. coli DH5 $\alpha$***

*Sau*3A1 fragments of *Z. mobilis* chromosomal DNA were used to provide a heterologous promoter that would not be subject to native regulatory mechanisms in *K. oxytoca* or interfere with subsequent integration into the *K. oxytoca* chromosome (Zhou, et al. (1999) *J. Indust. Microbiol. Biotechnol.* 22:600-607). Fragments of 0.5-1.5 kbp were isolated and randomly ligated into the *Bam*HI site of pLOI2317 to generate a library of surrogate promoters (See Figure 13). Approximately 75,000 colonies were screened on endoglucanase indicator plates. One-third of the clones actively produced *celY*. The most active 100 colonies were identified by zone size, purified, and re-tested. The 30 clones with the largest zones of activity were grown overnight in LB and assayed for CMCCase activity. The five most active are listed in Table 14 below, and exhibited approximately 7-fold higher activity than the original clone, pMH18. Plasmid pLOI2323 was selected for further investigation.

**TABLE 14. Expression of celY in DH5 $\alpha$  using fragments using Sau3A1 digestion products of Z. mobilis chromosomal DNA as surrogate promoters.**

Plasmids expressing celY or celZ	Endoglucanase activity (IU/L)		
	Extracellular	Total	% Extracellular
pMH18 (native celY promoter)	151	184	82
pLOI2317 (promoterless celY vector)	0	0	0
<i>celY</i> expressed from surrogate promoters			
pLOI2318	1,123	1,257	89
pLOI2319	888	1,023	87
pLOI2320	1,023	1,056	97
pLOI2323	1,257	1,291	97
pLOI2342	1,224	1,257	97
pLOI2349 (celZ)	3,414	16,234	21

5 All plasmids are pUC derivatives. Endoglucanase activity was measured using cultures grown at 37°C for approximately 16 hours.

The *Z. mobilis* Sau3A1 fragment (937 bp) in pLOI2323 was sequenced in both directions (GenBank Accession No. AF305919). Based on a database comparison, this 10 fragment appears to be derived from two pieces, a 882 bp fragment form *Z. mobilis* chromosome which corresponds to a previously sequenced region and a 55 bp fragment from the vector. A BLAST search (National Center for Biotechnology Information; <http://www.ncbi.nlm.nih.gov/BLAST/>) of the translated sequence did not reveal identity to known genes. Four sites of transcriptional initiation were identified in DH5 $\alpha$ .  
 15 (pLOI2323) by primer extension analysis involving three different sigma factors,  $\delta^{32}$ ,  $\delta^{38}$ , and  $\delta^{70}$  (See Figure 14). Although the differences in intensity were less than 2-fold, the sequence upstream from the most intense start site resembled the consensus for  $\delta^{32}$  (*rpoH*), the heat shock promoter (Wang, et al. (1998) *J. Bacteriology* 180:5626-5631; Wise, et al. (1996) *J. Bacteriology* 178:2785-2793).

20

***Construction of a vector for the integration of celY and celZ into the chromosome of K. oxytoca P2***

The plasmid pLOI2307 (7.2 kbp) was constructed and used to express celZ from a surrogate *Z. mobilis* promoter at high levels in recombinant *E. coli* DH5 $\alpha$  (See Zhou, et al. (1999) *B.Appl. Environ. Microbiol.* 65:2439-2445) and *K. oxytoca* M5A1 (See Zhou, et al. (1999) *Indust. Microbiol. Biotechnol.* 22:600-607). To facilitate subcloning of this hybrid celZ gene and promoter (4.5 kbp), an EcoRI linker was inserted into the T4 polymerase-treated *Sph*I site of pLOI2307 to provide flanking EcoRI sites for

convenient excision (pLOI2349). Prior to constructing a plasmid containing *celY* and *celZ*, a random 3 kbp fragment of EcoRI-digested *K. oxytoca* M5A1 chromosomal DNA was inserted into pLOI2323 containing *celY* (and surrogate promoter) to serve as a guide for homologous recombination (pLOI2348; 8 kbp). This 3 kbp M5A1 fragment was  
5 partially sequenced and appears to encode the complete M5A1 *glgP* gene. In pLOI2348 (8 kbp), flanking *AscI* sites allowed the excision of a single 5.5 kbp fragment containing the M5A1 *glgP* gene, *Z. mobilis* surrogate promoter, and *E. chrysanthemi* *celY*.

Figure 15 summarizes the construction of the *celY*, *celZ* integration vector from pLOI2349, pLOI2224, and pLOI2348. The recombinant *celY* and *celZ* genes containing 10 surrogate promoters and the guide fragment were sequentially inserted into the core integration vector, pLOI2224 (Martinez-Moralez, *et al.*, *supra*) using *E. coli* S17-1 as the host, to produce pLOI2352 (12 kbp). The 4.5 kbp EcoRI fragment from pLOI2349 containing *celZ* was inserted into pLOI2224 using an EcoRI site to make pLOI2350 (6.4 kbp). The 5.5 kbp *AscI* fragment from pLOI2348 containing *celY* was inserted into the 15 *AscI* site of pLOI2350 to make pLOI2352 (12 kbp). The fragments containing *cel* genes were oriented such that expression from the surrogate promoters was divergent. The resulting vector contained a R6K replicon that does not function in DH5 $\alpha$  or M5A1. The two FRT sites in pLOI2352 facilitate removal of the kanamycin gene and replicon after integration (Martinez-Moralez, *et al.*, *supra*).  
20

#### ***Functional integration of celY and celZ into the K. oxytoca P2 chromosome***

Plasmid pLOI2352 was introduced into P2 by electroporation followed by selection for kanamycin resistance. Approximately 150 colonies were recovered and all were positive on endoglucanase indicator plates. Ten clones with the largest zones of 25 activity were purified, grown in broth and assayed for endoglucanase activity. These produced 5-6 IU/ml of endoglucanase activity. One clone was selected for further study and designated as SZ12.

Due to the natural resistance of *K. oxytoca* to ampicillin, an additional antibiotic resistance marker (*tet*) was added to pFT-A plasmid containing the *flp* recombinase to 30 facilitate selection. The tetracycline gene was isolated as a 1.4 kbp EcoRI to *Aval* fragment from pBR322. After treatment with Klenow polymerase, this fragment was ligated into the Klenow-treated *ClaI* site of pFT-A to produce pLOI2353 (7.0 kbp). This plasmid encodes resistance to both ampicillin and tetracycline, the FLP recombinase (*flp*) under the control of the tetracycline promoter, and a temperature-conditional 35 pSC101 replicon.

Plasmid pLOI2353 was transformed into SZ12 and plated at 30°C with selection for tetracycline resistance. The presence of tetracycline also induced *flp* expression

resulting in a deletion of the kanamycin gene and R6K replicon from chromosomally integrated pLOI2353. Of 307 tetracycline-resistant colonies tested, >99% retained expression of the endoglucanase genes and were sensitive to kanamycin. Clones were purified, grown in broth and assayed for endoglucanase activity. All were similar and 5 one was designated SZ21(pLOI2353). The helper plasmid was eliminated from SZ21 by overnight growth at 37°C.

#### ***Construction of a celZ knockout mutation***

To confirm the presence of functional *celY* in SZ21, a knockout mutation of the 10 chromosomally integrated *celZ* was constructed by double, homologous recombination using plasmid pLOI2357 (Figure 16). Plasmid, pUC19 was digested with *Sma*I and *Hind*III, treated with Klenow polymerase, and self-ligated to eliminate many of the polylinker sites (pLOI2354). The remaining *Eco*RI site was used to insert a 4.5 kbp *Eco*RI fragment containing the promoter and *celZ* gene from pLOI2349 to make 15 pLOI2355 (7.2 kbp). The 1.8 kbp *Sma*I fragment from pHPO45aac containing the apramycin resistance gene (*aac*) was then ligated into the central region of *celZ*, replacing a small internal *Pst*I fragment (after blunting with T4 polymerase) to produce pLOI2356 (9.0 kbp). The 6.3 kbp *Eco*RI fragment from this plasmid was isolated and inserted into the core integration vector, pLOI2224, to produce pLOI2357 (8.2 kbp). 20 This plasmid contains a conditional R6K replicon and kanamycin resistance gene in addition to a *celZ* gene that is interrupted by an apramycin resistance gene.

Plasmid pLOI2357 was electroporated into SZ21 with selection for apramycin. Approximately 10% of the recombinants were apramycin resistant and kanamycin sensitive indicating a double homologous recombination event. These clones exhibited 25 low levels of endoglucanase production on indicator plates (See Table 15, below). One was selected and designated SZ22. Loss of EGZ and retention of EGY in SZ22 was confirmed by SDS-PAGE using the Pharmacia Phast Gel system.

It is interesting to note that cell clumping in liquid culture, typical of M5A1 and P2, was eliminated by the functional expression of *celZ* from integrated genes or from 30 plasmids. Clumping was not affected by the functional expression of *celY* alone.

#### ***Transcriptional initiation in K. oxytoca SZ21***

Primer extension analysis of *celY* and *celZ* in SZ21 were similar to those observed in DH5 $\alpha$ . A single major transcriptional start was identified for *celZ* that 35 corresponded precisely to the most prominent start site in DH5 $\alpha$  (pLOI2183) which contains the same promoter fragment (Zhou, et al. (1999) *J. Indust. Microbiol. Biotechnol.* 22:600-607; Zhou, et al. (1999) *B. Appl. Environ. Microbiol.* 65:2439-

2445). DNA immediately upstream from this site resembles the recognition sequence for a  $\sigma^{70}$  promoter (Wang, *et al., supra* and Wise, *et al., supra*). As observed with DH5 $\alpha$  (pLOI2323) (See Figure 14), primer extension analysis of *celY* indicated the presence of multiple putative transcriptional starts in SZ21. Although localized in the same regions  
5 as the start sites in DH5 $\alpha$ (pLOI2323), all bands were of near equal intensities.

***Effect of the E. chrysanthemi out genes (pCPP2006) on the extracellular secretion of EGY and EGZ in derivatives of K. oxytoca P2***

Table 7 summarizes the endoglucanase activities exhibited by cellulolytic  
10 derivatives of ethanologenic *K. oxytoca* P2. Strain SZ6 (Zhou, *et al.*, (1999) *J. Indust. Microbiol. Biotechnol.* 22:600-607) contains a chromosomally integrated hybrid *celZ* gene with same promoter fragment used to construct SZ21. Despite the presence of two  
15 endoglucanase genes in SZ21, extracellular and total endoglucanase activities were 13% lower in this strain than in SZ6. Most of the endoglucanase activity produced by SZ21 can attributed to *celZ*. SZ22, a *celZ* mutant of SZ21, expressed only 11% of the endoglucanase produced by the parent containing functional *celY* and *celZ* genes. In strains SZ6 and SZ21 containing a functional *celZ*, most of the endoglucanase activity (primarily EGZ) was cell associated. In strain SZ22 containing a functional *celY* alone, half of the endoglucanase activity was extracellular.  
20

**TABLE 15. Effect of out genes (pCPP2006) on endoglucanase production by derivatives of *K. oxytoca* P2.**

Strains	CMC zone (mm)	OD <sub>550</sub>	CMCase Activity (IU/L) <sup>a</sup>		
			Extracellular	Total	Secretion (%)
P2	0	10.5	0	0	0
SZ6	8.5	11.0	1,920	8,800	22
SZ21	6.7	11.0	1,620	7,800	21
SZ22	2.0	10.0	480	879	55
P2 (pCPP2006)	0	10.0	0	0	0
SZ6 (pCPP2006)	10.8	9.6	13,800	22,300	62
SZ21 (pCPP2006)	11.5	10.2	20,100	26,900	75
Spezyme CE <sup>®</sup> (10 ml/liter) <sup>b</sup>	-	-	-	27,000	-
Spezyme CP <sup>®</sup> (10 ml/liter) <sup>b</sup>	-	-	-	33,400	-

25     <sup>a</sup> Endoglucanase activity was measured using cultures grown at 30°C in LB containing 5% sorbitol for 24 h.

<sup>b</sup> Dilution equivalent to the highest Spezyme<sup>®</sup> level used in fermentation experiments (Table 4).

- The addition of the *out* genes (pCPP2006) to recombinant *E. coli* and *K. oxytoca* M5A1 harboring *celZ* can cause a dramatic increase in the functional expression of *celZ* and in the fraction of EGZ secreted into the extracellular milieu (Zhou, *et al.*, (1999) *J. Indust. Microbiol. Biotechnol.* 22:600-607; Zhou, *et al.* (1999) *B. Appl. Environ. Microbiol.* 65:2439-2445). These same effects were observed for ethanologenic *K. oxytoca* SZ21 containing *celZ* and *celY* (See Table 14). Addition of the *out* genes to SZ22 (inactive *celZ*) had no effect on the functional expression of *celY* or the extent of EGY secretion. This *celZ* mutant, SZ22 (pCPP2006), produced only 3% of the total endoglucanase activity (EGY) produced by SZ21 (pCPP2006) containing functional *celY* and *celZ* genes. The secreted endoglucanase produced by SZ21 with the *out* genes was substantially higher than the sum of the individual activities expressed from the same respective promoters in SZ6 (EGZ) and SZ22 (EGY), consistent with synergy between these two enzymes. In this assay, synergy is estimated to be 1.4-fold the arithmetic sum of the individual activities (SZ6(pCPP2006) and SZ22(pCPP2006)) for the combination of extracellular enzymes produced by SZ21(pCPP2006).

***Synergism between recombinant E. chrysanthemi endoglucanase (EGZ and EGY) and fungal cellulase (Spezyme CE®) during the fermentation of cellulose to ethanol***

- SSF tests in flasks were used without pH control to evaluate the combined effects of fungal cellulase (Spezyme®) and cellulase enzymes produced by the biocatalysts on ethanol production from Sigmacell 50®, a highly crystalline substrate. Results are indicated in Table 16, below.

**TABLE 16. Maximum ethanol production**

25

Strain	Genencor Spezyme®		Fermentation <sup>a)</sup>		
	Type	Addition (ml/liter)	N	Ethanol ± SD (g/liter) <sup>b)</sup>	% Control ± SD
P2 (pCPP2006)	none	0	3	0.23 ± 0.01	100 ± 2.0
SZ6(pCPP2006)	none	0	3	0.28 ± 0.02 *	124 ± 8.5
SZ21(pCPP2006)	none	0	3	0.26 ± 0.02 *	116 ± 8.5
SZ22(pCPP2006)	none	0	3	0.24 ± 0.01	107 ± 1.0
P2(pCPP2006)	CE	5.0	6	13.7 ± 0.3	100 ± 2.0
SZ6(pCPP2006)	CE	5.0	6	13.8 ± 0.3	101 ± 2.4
SZ21(pCPP2006)	CE	5.0	6	16.0 ± 0.5 **	117 ± 3.5
SZ22(pCPP2006)	CE	5.0	6	15.2 ± 0.3 **	112 ± 1.5
P2(pCPP2006)	CE	10.0	6	20.7 ± 0.5	100 ± 2.1
SZ6(pCPP2006)	CE	10.0	6	21.2 ± 0.1	103.4 ± 0.4

**TABLE 16. Maximum ethanol production (Continued)**

SZ21(pCPP2006)	CE	10.0	6	$24.6 \pm 0.5$ **	$121 \pm 2.3$
SZ22(pCPP2006)	CE	10.0	6	$25.2 \pm 1.1$ **	$122 \pm 5.0$
P2(pCPP2006)	CP	5.0	6	$15.2 \pm 0.3$	$100 \pm 1.7$
SZ21(pCPP2006)	CP	5.0	6	$17.8 \pm 0.4$ **	$116 \pm 2.2$
P2(pCPP2006)	CP	10.0	6	$25.3 \pm 0.7$	$100 \pm 2.6$
SZ21(pCPP2006)	CP	10.0	6	$27.2 \pm 0.3$ **	$107 \pm 1.2$

<sup>a)</sup> Cultures without added cellulose (100 g/liter) or Spezyme® produced  $0.22 \pm 0.01$  g/liter of ethanol. Spezyme® contained approximately 100 FPU/ml. Addition of 5 ml and 10 ml of Spezyme® corresponds to 5 FPU/g and 10 FPU/g cellulose, respectively.

<sup>b)</sup> Student t-test shows that there is significant difference in ethanol production compared to the respective P2 controls at each Spezyme® dilution. P value ≤0.001 is indicated by a two asterisks.  
P value ≤0.05 is indicated by a single asterisk.

\* P value  $\leq 0.05$  is indicated by a single asterisk.

10

Although very low levels of ethanol were produced by all strains in the absence of Spezyme®, strains SZ6(pCPP2006) and SZ21(pCPP2006) containing functional *celZ* genes produced higher levels of ethanol ( $p \leq 0.05$ ) than strain SZ22(pCPP2006) containing only a functional *celY* gene and strain P2(pCPP2006) lacking endoglucanase genes. In the absence of both Spezyme® and Sigmacell 50®, all strains produced 0.22 g/L ethanol. The additional increment of ethanol produced by SZ6(pCPP2006) and SZ21(pCPP2006) during incubation with Sigmacell 50 is attributed to hydrolysis of the small fraction of amorphous cellulose in the substrate by EGZ (Zhou, et al. (1999) *J. of Industrial Microbiol. Biotechnol.* 22:600-607; Zhou, et al. (1999) *B. Appl. Environ.*

20 *Microbiol.* 65:2439-2445; Zhou, S., et al. (2000) *J. Bacteriol.* 182:5676-5682). Digestion of amorphous cellulose by EGY produces saccharides that are too large to be transported and metabolized without further hydrolysis (Zhou, S., et al. (2000) *J. Bacteriol.* 182:5676-5682).

Spezyme CE® and Spezyme CP® contain a commercially optimized combination of endoglucanase, exoglucanase, and cellobiase activities (Beguin, *et al.* (1994) *FEMS Microbiol. Rev.* 13:25-58; Boyer, *et al.* (1987) *Eur. J. Biochem.* 162:311-316; Nieves, *et al.* (1998) *World Journal of Microbiology and Biotechnology* 14:310-314; Ohmiya, *et al.* (1997) *Biotechnol. Genetic Eng. Rev.* 14:365-414). Despite this optimization, Spezyme®-supplemented fermentations with two of the endoglucanase producing biocatalysts, SZ21(pCPP2006) and SZ22(pCPP2006), produced significantly higher levels of ethanol than the control P2(pCPP2006) which lacks endoglucanase genes. The combinations of Spezyme® and SZ21(pCPP2006) and SZ22(pCPP2006) were synergistic in terms of ethanol production, up to 20% higher than the sum of ethanol produced by each individually ( $p \leq .005$ ). Synergy was observed for both dilutions of Spezyme CE®, and for Spezyme CP®. This synergistic effect can be attributed primarily

to EGY since this is the only endoglucanase produced by SZ22(pCPP2006). No synergy was observed for SZ6(pCPP2006) which produces only EGZ.

## EXAMPLE 5

5      **Simultaneous Saccharification and Fermentation of Amorphous Cellulose to Ethanol by Recombinant *Klebsiella oxytoca* SZ21 without Supplemental Cellulase**

In this example, a derivative of *Klebsiella oxytoca* M5A1 containing chromosomally integrated genes for ethanol production from *Zymomonas mobilis* (*pdc*, *adhB*) and endoglucanase genes from *Erwinia chrysanthemi* (*celY*, *celZ*) is demonstrated 10 to produce over 20,000 U/L of endoglucanase activity during fermentation. In particular, this strain is demonstrated to, in combination with its native ability to metabolize cellobiose and cellooligosaccharides, ferment amorphous cellulose to ethanol (58-76% of theoretical yield) without the addition of cellulase enzymes from other organisms.

15      Throughout this example, the following materials and methods are used unless otherwise stated.

### ***Materials and Methods***

#### *Bacteria, plasmids and culture conditions*

Four ethanologenic derivatives of *Klebsiella oxytoca* M5A1 were used in this 20 study. Strain P2 contains chromosomally integrated *pdc* and *adhB* genes from *Z. mobilis* for ethanol production (Wood *et al.*, (1992) *Appl. Environ. Microbiol.* 58: 2103-2110). This strain was the parent organism for three strains that contain highly expressed, chromosomally integrated endoglucanase genes from *E. chrysanthemi* (Zhou *et al.* 25 (2001) *Appl. Environ. Microbiol.* 67: 6-14). Strains SZ6, SZ22, and SZ21 contain *celZ* alone, *celY* alone, and both *celY* and *celZ*, respectively. Additional genes (approximately 15 *out* genes) were required for the efficient secretion of endoglucanase CelZ (He *et al.* (1991) *Proc. Natl. Acad. Sci. USA* 88:1079-1083; Pugsley *et al.*, (1997) *Gene* 192: 13-19).and were supplied by plasmid pCPP2006. For consistency, this plasmid has been inserted into all strains. Recombinant *K. oxytoca* strains were grown 30 in Luria broth (per liter: 10 g yeast extract, 5 g tryptone, 5 g NaCl) containing a carbohydrate. Spectinomycin (100 mg/L) was included to maintain plasmid pCPP2006. Seed cultures (5% glucose, 12-16 h) and fermentations were incubated at 35°C (120 rpm) and were vented using a 22 gauge needle.

35      *Assay of endoglucanase activity*

Total endoglucanase activity (Cel Y and CelZ) was determined for cultures grown for 24 hours in Luria broth containing 5% sorbitol to minimize interference

during the determination of reducing sugars (Zhou et al. (2000) *J. Bacteriol.* 182: 5676-5682). Cell suspensions in broth were briefly treated with ultrasound to release bound enzymes. Dilutions were assayed at 35°C in 50 mM sodium citrate buffer (pH 5.2) using carboxymethyl cellulose as the substrate (2%). Reactions were terminated by 5 heating (10 min, 100°C). Activity was expressed as µmole per min of reducing sugar (U) using glucose as a standard.

*Thin layer chromatography (TLC) analysis of cellobiosides*

Cellobiosides were separated using Whatman LK5 silica gel plates (Cat. No. 10 4855-620). Plates were developed in solvent (6 parts chloroform, 7 parts acetic acid, and 1 part water), air-dried, and developed a second time in the same solvent. Saccharides were visualized by spraying with 6.5 mM N-(1-naphthyl)ethylenediamine dihydrochloride and heating to 100°C for 10 min (Zhou et al. *supra*).

15 *Preparation of cellobiosides from microcrystalline cellulose*

Cellulose was converted to cellobiosides based on the method of Pereira et al. (1988). Sigmacell type 50 cellulose (100 g; Cat. No. S-5504) was slowly added to 400 g of 72% H<sub>2</sub>SO<sub>4</sub> (w/w) and allowed to hydrolyze at 22°C with stirring for 18 hr. The digested slurry was precipitated by slowly adding to 2.5 L of cold ethanol (100%), 20 centrifuged, and the pellet washed twice with 1 L of cold ethanol. The pellet was dissolved in 100 ml of H<sub>2</sub>O, adjusted to pH 6, and centrifuged at 5000 x g for 20 min to remove insoluble fibers. The supernatant containing cellobiosides was dried and analyzed.

25 *HPLC analysis of cellobiosides*

Cellobiosides were analyze using a Waters HPLC equipped with a refractive index monitor and digital integrator. Separations were performed at 65°C using a BioRad HPX-42A column with distilled water as the mobile phase (0.6 ml/min).

30 *Fermentation of cellobiosides*

Initial fermentations used cellobiosides purchased from the Sigma Chemical Company (Cat. No. C- 8071). Seed cultures (112.5 µL) were typically combined with 37.5 µL of 4% cellobiosides in a 1.5-ml microcentrifuge tube and incubated for 36 h (35°C and 120 rpm). Samples (50 µL each; 0 h, 10 h, and 36 h) were heated (100°C, 10 35 min) to inactivate enzymes and centrifuged (10,000 x g, 5 min) to remove cell debris. Supernatants (4 µL) were then spotted on TLC plates for analysis.

Small fermentations (50-ml flasks, 5 ml broth volume) were conducted in a similar fashion with laboratory samples of cellobiosides prepared by sulfuric acid hydrolysis. Cellobiosides (6 g/L total for saccharides less than 7 glucosyl residues) were filter sterilized in Luria broth. Flasks were inoculated with pelleted cells (5,000 x g, 10 min) from a seed culture (dry weight of 0.33 g/L) to eliminate ethanol produced from the glucose in seed media. Samples were removed initially, and then at 24-h intervals for the determination of the presence of ethanol by gas chromatography (Beall *et al.* (1991) *Biotechnol Bioeng*. 38: 296-303).

10 *Preparation of amorphous cellulose*

Phosphoric acid-swollen cellulose was prepared by a modification of the method described by Wood (1988). Approximately 20 g of Avicel (Cat. No. PH105, Fluka Chemika) was added slowly to 500 ml 85% H<sub>3</sub>PO<sub>4</sub>, stirred at room temperature overnight, poured into 4 L of ice cold water, and allowed to settle without agitation for 15 min. After decanting the upper liquid layer, 4 L of cold water was added and mixed thoroughly, and repeated 5 times with water, once with 1% NaHCO<sub>3</sub>, and 5 more times with water (final pH 6-7). The cellulose suspension was concentrated by centrifugation (5000 x g, 20 min) and used as a substrate for fermentation. This viscous suspension contained a mixture of crystalline and amorphous cellulose.

20 The fraction of amorphous cellulose present in the phosphoric acid-swollen cellulose was estimated by repeated digestion with CelY and CelZ endoglucanases. A 24-h culture of SZ21(pCPP2006) was sonicated briefly, centrifuged (5,000 x g, 10 min), and the supernatant used as a source of endoglucanase (~20 U/ml). Approximately 1 g of the viscous, acid-swollen cellulose was weighed into each of 6 pre-weighed 25 centrifuge tubes. Two served as controls (no enzyme) and were dried to determine initial dry weight and moisture content. Endoglucanase (9 ml) was added to 4 tubes and these were incubated at 35°C for 12 h. Chloroform (2 drops) was added to each tube to retard microbial growth. After centrifugation (5,000 x g, 20 min), this process was 30 repeated with new enzyme preparations for a total of 6 successive treatments over a 72 h period. Tubes were removed at various times, centrifuged, washed once with distilled water to remove soluble products and salts, and dried to a constant weight at 70°C. Amorphous cellulose was calculated as the reduction in dry matter resulting from endoglucanase digestion.

35 *Fermentation of amorphous cellulose*

Acid swollen cellulose (40 g, unsterilized) was combined with 5 ml of a 10X concentrate of Luria broth and 5 ml of seed culture (cell dry weight of 0.33 mg/L) in a

125-ml flask. Both spectinomycin (100 µg/ml) and chloramphenicol (40 µg/ml) were added to eliminate contamination. The viscous mixture was initially mixed using an applicator stick, then incubated at 35°C. Due to the high initial viscosity, flasks were mixed at 200 rpm during the first 2 h and subsequently, at 120 rpm. Ethanol was  
5 measured by gas chromatography (Beall *et al.* (1991) *Biotechnol. Bioeng.* 38:296-303).

### *Viscosity*

The viscosity of acid-swollen cellulose preparations was estimated at 22°C by timing the flow through a vertical 10-ml glass pipette. Solutions of glycerol were used  
10 as standards. Flow times of 2 sec to 75 sec were observed for fermentation broths, corresponding to viscosities of 1 to 1,300 centipoise, respectively.

### **Results**

#### **Comparison of endoglucanase activities produced by ethanologenic derivatives of *K. oxytoca***

Ethanologenic strains containing endoglucanase genes from *E. chrysanthemi* produced substantial levels of activity during glucose fermentation. The highest activity was produced by strain SZ21(pCPP2006) containing both *celZ* and *celY*,  $29.3 \pm 1.6$  U/ml of culture. Strain SZ6(pCPP2006) containing *celZ* alone produced  $22.5 \pm 1.7$  U/ml  
20 and strain SZ22(pCPP2006) containing *celY* (*celZ* deletion of strain SZ21) produced  $1.0 \pm 0.1$  U/ml. Approximately 60% of the activity was secreted by each strain, the balance being cell associated and readily released by mild sonication. No endoglucanase activity was detected in the parent, strain P2(pCPP2006).

#### **25 Analysis of cellobiosides**

The presence of cellobiosides was analyzed by thin layer chromatography and by HPLC. Preparations from the Sigma Chemical Company and those prepared in our laboratory were similar. Lane 1 of both thin layer chromatograms (Figure 17, panels A and B ) shows a representative separation of the Sigma product at the beginning of  
30 fermentation. Small amounts of glucose and cellobiose were present with intermediate amounts of celotriose and cellohexaose, and larger amounts of cellotetraose and cellopentaose. The most intense region, however, remained at the origin. This region contains cellobiosides of greater than 6 residues and other uncharacterized compounds.

HPLC analyses were in agreement with thin layer chromatograms. The Sigma  
35 product contained 3% glucose, 3% cellobiose, 13% celotriose, 25% cellotetraose, 22% cellopentaose, and 9% cellohexaose. Based on peak areas, approximately 25% of the saccharides were longer than 6 residues. Cellobiosides prepared in our laboratory

contained 7% glucose, 6% cellobiose, 10% cellotriose, 15% cellobetraose, 12% cellopentaose, 19% cellohexaose, and 30% saccharides longer than 6 glucosyl residues.

#### ***Fermentation of cellobiosides***

- 5       Figure 17 shows a thin layer chromatogram comparing the fermentation of cellobiosides (Sigma Chemical Company) by strain P2(pCPP2006) and by SZ21(pCPP2006), a derivative which secretes both endoglucanases CelY and CelZ. Utilization of glucose, cellobiose, and cellotriose by strain P2 was confirmed and is particularly evident after 36 h. However, this strain was unable to metabolize longer  
10 cellobiosides. In contrast, strain SZ21 degraded and metabolized virtually all of the separated cellobiosides and a portion of the material at the origin. It is relevant to note that after 10 h, the levels of cellobiose and cellotriose in the SZ21 fermentation were several-fold higher than initially present. This increase was attributed to the hydrolysis of longer cellobiosides by secreted CelY and CelZ.
- 15       The effectiveness of secreted endoglucanases was also examined during ethanol production from laboratory cellobiosides (Table 17). Although only low levels of ethanol were produced due to the low substrate concentrations, the benefit of CelZ endoglucanase is clearly evident. The parent P2(pCPP2006) and strain SZ22(pCPP2006; CelY only) produced half as much ethanol as the two strains that secreted CelZ:  
20 SZ6(pCPP2006; CelZ only) and SZ21(pCPP2006; CelY and CelZ). Endoglucanase CelY was of no benefit using cellobiosides as substrates, consistent with the requirement for a long-chain substrate.

- 25       Ethanol yields were estimated based on cellobiosides containing less than 7 glucosyl residues (Table 17). An average of 62% of the theoretical yield was observed for the two strains producing endoglucanase CelZ. Higher yields (over 90%) were previously measured during the fermentation of 100 g/L cellobiose by strain P2 (Wood & Ingram 1992). It is likely that the lower yields observed with low concentrations of cellobiosides result in part from evaporative losses and the diversion of a larger fraction of carbon to cell growth.  
30

#### ***Amorphous cellulose content of phosphoric acid swollen cellulose***

- 35       Acid swollen cellulose partially recrystallizes during the removal of acid and storage due to extensive hydrogen bonding between cellulose ribbons. *E. chrysanthemi* Cel Y and Cel Z hydrolyze amorphous cellulose and carboxymethyl cellulose but are unable to hydrolyze crystalline cellulose. This resistance of crystalline cellulose to endoglucanase hydrolysis was used to estimate the fraction of washed, acid-swollen cellulose which remained amorphous as the loss of insoluble material (dry weight).

Samples of amorphous cellulose were repeatedly incubated with fresh endoglucanase preparations (~20 U/ml). After an initial 12-h treatment, 28% of the dry weight was solubilized. After four successive treatments, 43% was solubilized and remained constant for the two final treatments. This value, 43%, represents an estimate of the 5 fraction of cellulose that remained in the amorphous state. The actual fraction of amorphous cellulose may be somewhat lower since estimates also include losses of crystalline cellulose which may have occurred during centrifugation and washing.

### ***Fermentation of amorphous cellulose***

10 The functional expression of high levels of endoglucanase is sufficient to permit the direct conversion of amorphous cellulose to ethanol by two derivatives of *K. oxytoca* P2 (Table 17; Figure 18). Three concentrations of amorphous cellulose were tested with similar results.

15 The highest levels were produced by strain SZ21(pCPP2006) which secretes a combination of CelY and CelZ endoglucanases. Lower levels of ethanol were produced by strain SZ6(pCPP2006) which secretes CelZ alone. A small amount of ethanol was produced from residual glucose in the seed media by the parent strain, P2(pCPP2006; no endoglucanase) and by strain SZ22(pCPP2006) which secreted only CelY. The combined effects of CelY and CelZ appears to be synergistic for ethanol production and 20 ethanol yield despite the lack of efficacy of CelY alone. Ethanol yields for strain SZ21(pCPP2006; both enzymes) ranged from 58%-76% of the theoretical maximum. This synergistic effect on ethanol production may result from increased hydrolysis due to differences in substrate specificity.

### ***Viscosity changes during the fermentation of amorphous cellulose***

Pronounced viscosity changes were observed during the fermentation of amorphous cellulose by endoglucanase-secreting strains (Figure 18). The largest reduction in viscosity was evident during the first 2 h of incubation for SZ6(pCPP2006; CelZ alone) and SZ21(pCPP2006; CelY and CelZ), from a viscosity near that of pure 30 glycerol to that near water. Less pronounced changes were observed with SZ22(pCPP2006; CelY alone) and no change was evident with the parent strain P2(pCPP2006) which served as a control. Changes in viscosity were essentially complete after 12 h.

Differences in viscosity were quantified at the end of fermentation for 35 Experiment 4 in Table 17 (Figure 18, panel C). These varied by three orders of magnitude. The highest final viscosity was observed for P2(pCPP2006) which does not produce endoglucanase, 1,300 centipoise. Viscosity was reduced by half (500

centipoise) during fermentation with SZ22(pCPP2006) as a result of hydrolysis by endoglucanase CelY alone. At the end of fermentation, the viscosities of broths from both strains that secrete CelZ were essentially equivalent to that of water, 1 centipoise for SZ21(pCPP2006; CelY and CelZ) and 2 centipoise for SZ6(pCPP2006; CelZ alone).

- 5 Clearly, CelZ alone was more effective than CelY alone. No further benefit was attributed to the production of CelY in combination with CelZ when compared at the end of fermentation (96 h) although both enzymes function synergistically in the production of ethanol (Table 17; Figure 18) and during saccharification.

In summary, these results using, e.g., *K. oxytoca* strain SZ21, demonstrate an  
10 advancement toward the goal of producing sufficient cellulase enzymes for the direct bioconversion of cellobiosides and amorphous cellulose to ethanol without the addition of supplemental enzymes. Endoglucanase levels produced by this strain are over 10-fold over those previously reported for engineered strains of yeast and other bacteria during ethanol fermentation (Brestic-Goachet *et al.* 1989, Cho *et al.* 1999, Cho & Yoo  
15 1999, Misawa *et al.* 1988, Su *et al.* 1993, Van Rensburg *et al.* 1996, 1998). Moreover, the level of endoglucanase produced by strain SZ21 is roughly equivalent to 1% of the endoglucanase activity present in commercial cellulase concentrates (Nieves *et al.* 1998, Tomme *et al.* 1995, Wilson *et al.* 1997).

The effectiveness of strain SZ21 in the hydrolysis of amorphous cellulose results  
20 from the combination of two endoglucanase enzymes *E. chrysanthemi* and function synergistically (Guiseppi *et al.* 1991, Zhou & Ingram 2000). Moreover, these secreted enzymes function with the phosphotransferase system in *K. oxytoca* which provides efficient uptake and metabolism of cellobiose and cellooligosaccharides (Wood & Ingram 1992, Lai *et al.* 1997).

**TABLE 17. Production of ethanol from cellobiosides and from amorphous cellulose**

Expt.	Strain	Substrate	Substrate (g/L) <sup>(a)</sup>	N	Ethanol <sup>(b)</sup> (g/liter)	% Theoretical Yield <sup>(c)</sup>
1	P2 (pCPP2006)	cellobiosides <sup>(d)</sup>	6.0	2	1.06	33
1	SZ6(pCPP2006)	cellobiosides	6.0	2	2.04	63
1	SZ21(pCPP2006)	cellobiosides	6.0	2	2.02	62
1	SZ22(pCPP2006)	cellobiosides	6.0	2	0.93	26
2	P2(pCPP2006)	amorphous cellulose	6.85	3	1.77 ± 0.06	0
2	SZ6(pCPP2006)	amorphous cellulose	6.85	3	3.97 ± 0.12**	57
2	SZ21(pCPP2006)	amorphous cellulose	6.85	3	4.67 ± 0.06**	76
2	SZ22(pCPP2006)	amorphous cellulose	6.85	3	1.80 ± 0.01	0
3	P2 (pCPP2006)	amorphous cellulose	15.34	3	1.85 ± 0.01	0
3	SZ6(pCPP2006)	amorphous cellulose	15.34	3	6.07 ± 0.18**	49
3	SZ21(pCPP2006)	amorphous cellulose	15.34	3	7.84 ± 0.35**	70
3	SZ22(pCPP2006)	amorphous cellulose	15.34	3	1.92 ± 0.02	0
4 <sup>(e)</sup>	P2 (pCPP2006)	amorphous cellulose	28.96	1	1.90	0
4	SZ6(pCPP2006)	amorphous cellulose	28.96	1	10.1**	51
4	SZ21(pCPP2006)	amorphous cellulose	28.96	1	11.3**	58
4	SZ22(pCPP2006)	amorphous cellulose	28.96	1	1.80	0

- 5 <sup>a</sup> Inocula for cellobioside fermentations were harvested by centrifugation to eliminate seed broth as a source of ethanol. An average of 1.85 g/L ethanol was produced from residual sugar in the seed broth used to inoculate the three fermentations of amorphous cellulose.  
<sup>b</sup> \*\* Indicates that the value is significantly different from that of the control strain (P2) lacking genes encoding *E. chrysanthemi* endoglucanases ( $p<0.001$ ).
- 10 <sup>c</sup> Yields are calculated as a percentage of theoretical maxima, assumed to be 0.54 g ethanol per gram of cellobiose and 0.56 g ethanol per gram of amorphous cellulose. For amorphous cellulose, yields were computed after subtracting ethanol produced from sugar in the inocula.  
<sup>d</sup> Cellobiosides (g/L) represents the sum of glucose plus cellobiosides containing fewer than 7 glucosyl residues.
- 15 <sup>e</sup> The highest coefficient of variation in Experiments No. 2 and No. 3 was 4.5% of the mean value. By assuming a similar coefficient of variation (5% of the measured value), significance was also estimated for Experiment No. 3.

*Equivalents*

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following  
5 claims. Moreover, any number of genetic constructs, host cells, and methods described in United States Patent Nos. 5,821,093; 5,482,846; 5,424,202; 5,028,539; 5,000,000; 5,487,989, 5,554,520, and 5,162,516, may be employed in carrying out the present invention and are hereby incorporated by reference.

U.S. GOVERNMENT USE